DOCUMENT RESUME

ED 336 109 IR 053 706

AUTHOR	Johnson, Margaret; And Others
TITLE	Computer Files and the Research Library.
INSTITUTION	Research Libraries Group, Inc., Mountain View, CA.
PUB DATE	Nov 90
NOTE	57p.; Proceedings of an Invitational Machine-Readable Data Files Workshop Stanford, CA, 1989). Edited by Constance C. Gould.
AVAILABLE FROM	The Research Libraries Group, Inc., Distribution Services Center, 1200 Villa Street, Mountain View, CA 94041-1100 ($3.50 plus sales tax).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Descriptive (141)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Academic Libraries; Access to Information; Cataloging; Classification; Computer Networks; *Databases; Higher Education; Information Retrieval; Library Collections; *Library Services; Policy Formation; *Records Management; *Research Libraries; Workshops
IDENTIFIERS	Research Libraries Information Network

ABSTRACT

This report on a machine-readable data files workshop highlights unresolved issues related to innovative approaches to collecting, describing, and providing service for computer files in research libraries. The remarks of four of the workshop speakers are presented. Margaret Johnson of the University of Minnesota gives a comprehensive review of collection management and development issues, from understanding the medium to training bibliographers. Anita Lowry of Columbia University introduces the Electronic Text Service, which she describes as probably the first research-library-based research and instructional facility devoted to machine-readable humanities texts. Lynn Marko of the University of Michigan describes her library's ground-breaking project to catalog Research Libraries Information Network (RLIN) data files from the Inter-university Consortium for Political and Social Research. Finally, Katherine Chiang of Cornell University presents a detailed checklist of the skills and expertise required to provide access to computer files, beginning with selecting the computer file and ending with assisting patrons in data extraction. Concluding the report are the workshop agenda, a list of the workshop attendees and their mailing addresses, a summary of the workshop discussion results, and summaries of six projects: (1) "Computer Files in the Research Library: Issues of Collection and Access" (Cornell University): (2) "Capturing Information about Machine-Readable Files in the Humanities" (Dartmouth College); (3) "Progress in MRDF: The Machine-Readable Data File Project" (University of Florida); (4) "Policy Development for the Acquisition of Machine-Readable Data: A Demonstration Project" (Northwestern University); (5) "Covering the Bases: Executive Summary" (New York University); and (6) "Providing Local Access to Machine-Readable Data Files: Choices and Tradeoffs" (University of Pennsylvania). (MAB)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

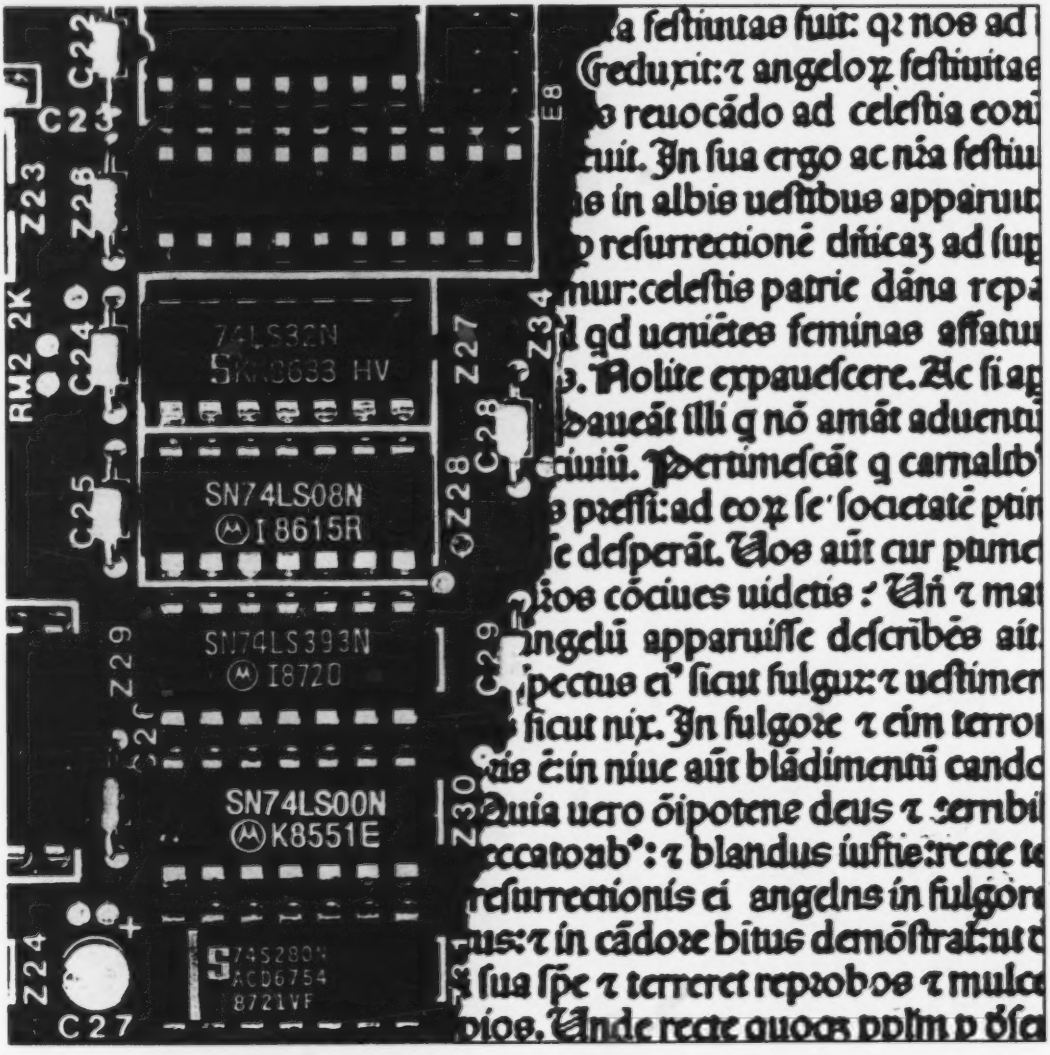
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Computer Files and the Research Library

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Jennifer Hartzell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



The Research Libraries Group, Inc. Mountain View, California, U.S.A.

1105370b

BEST COPY AVAILABLE

Computer Files and the Research Library

by Margaret Johnson, Anita Lowry,
Lynn Marko, and Katherine Chiang

Edited by Constance C. Gould

The Research Libraries Group, Inc.
Mountain View, California, U.S.A.

© 1990 by The Research Libraries Group, Inc.
All rights reserved

First published in November 1990
Printed in the United States of America

RLIN® is a registered service mark of
The Research Libraries Group, Inc.
Mountain View, California 94041-1100

Contents

Introduction.
By Constance C. Gould, Research Libraries Group, Inc. 1

Adding Computer Files to the Research Library:
Issues in Collection Management and Development.
By Margaret Johnson, St. Paul Campus Libraries,
University of Minnesota. 3

Machine-Readable Texts in the Academic Library:
The Electronic Text Service at Columbia University.
By Anita Lowry, Butler Reference Library, Columbia
University Libraries. 15

Bibliographic Description of Computer Files:
Lessons Learned from a Case Study.
By Lynn Marko, Hatcher Graduate Library,
University of Michigan. 25

Computer Files in the Library: Training Issues.
By Katherine Chiang, Mann Library, Cornell University. 31

Results of the Research Libraries Group, Inc., Workshop on
Machine-Readable Data Files. 37
- Agenda 39
- Workshop Attendees 40
- Workshop Discussion Results 42
  - Collection Management and Development 42
  - Bibliographic Control and Access 43
  - Public Services 45

Appendix: Project Summaries from *Managing a New*
Library Resource: Results of the Machine-Readable Data Project
in Six Member Libraries. 49
- Computer Files in the Research Library:
  Issues of Collection and Access *(Cornell University)* 51
- Capturing Information About Machine-Readable
  Files in the Humanities *(Dartmouth College)* 53
- Progress in MRDF: The Machine-Readable
  Data File Project *(University of Florida)* 54
- Policy Development for the Acquisition of
  Machine-Readable Data: A Demonstration Project
  (Northwestern University) 56
- "Covering the Bases" Executive Summary
  (New York University) 58
- Providing Local Access to Machine-Readable
  Data Files: Choices and Tradeoffs
  (University of Pennsylvania) 59

Introduction

In the fall of 1989 the Research Libraries Group, Inc., held an invitational machine-readable data files workshop on the Stanford University campus. The decision to hold this workshop grew out of the desire to build on the work initiated by RLG members participating in a machine-readable data file (MRDF) project and to shape an agenda for future initiatives, both within RLG program committees and in collaboration with institutions outside RLG.

The original MRDF project, begun in 1987, arose from a need that had become acutely evident to libraries by the mid-1980s: computer files* in all disciplines were rapidly growing in numbers, complexity, and importance to research, and libraries were finding the existing infrastructure inadequate for this demanding new format, regardless of the variety of data (bibliographic, textual, software) or media (diskette, tape, compact disk). Funding from the Pew Charitable Trusts enabled six RLG institutions (Dartmouth College, Cornell, New York, and Northwestern universities, and the universities of Florida and Pennsylvania) to develop model projects for collecting, controlling, and providing access to computer files. RLG published the final project reports in June 1989.**

The fall 1989 workshop allowed the RLG consortium to take the Pew Project a step beyond this important beginning. Participants were invited from a cross-section of the RLG membership—the Board of Governors and standing program committees—as well as from several organizations outside RLG working actively with computer files. The gathering combined focused presentations and brainstorming in small groups. Its purpose was to clarify issues and begin to set priorities in the computer file arena rather than to prescribe a detailed agenda. Although recognizing the pressing nature of individual institutional problems, workshop participants emphasized capturing the opportunities for cooperative approaches, where cooperative was not equated solely with the RLG membership. Summaries of the conclusions reached in the workshop discussions—organized by the topic areas of collection management, bibliographic control and access, and public services—include suggestions for cooperative action.

At the workshop, participants expressed a desire for a succinct publication describing innovative approaches to collecting, describing, and providing service for computer files in research libraries. The publication envisioned would also highlight issues as yet unresolved and help to point the way for other libraries.

In *Computer Files and the Research Library*, we have endeavored to provide such a publication. Margaret Johnson of the University of Minnesota gives a comprehensive review of collection management and development issues, from understanding the medium to training bibliographers.

---

*In keeping with Library of Congress practice, we have dropped the term *machine-readable data file*, or MRDF, in favor of *computer file*, a broader and more comprehensive term that includes databases and software as well as any type of data file.

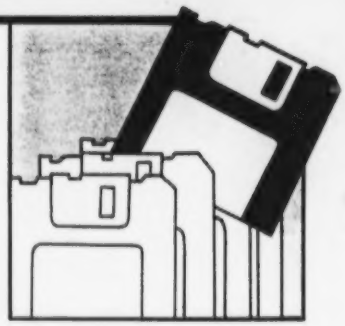
**_Managing A New Library Resource: Results of the Machine-Readable Data File Project in Six Member Libraries_. Mountain View, CA: Research Libraries Group, Inc., 1989. ($15.00).

Anita Lowry of Columbia University introduces us to the Electronic Text Service, probably the first research-library-based research and instructional facility devoted to machine-readable humanities texts. Lynn Marko of the University of Michigan describes her library's ground-breaking project to catalog RLIN® data files from the Inter-university Consortium for Political and Social Research. Finally, Katherine Chiang of Cornell University gives us a detailed checklist of the skills and expertise required to provide access to computer files, beginning with selecting the computer file and ending with assisting patrons in data extraction.

Since the workshop, RLG committees have made progress on a number of fronts. Members of the Public Services, Collection Management and Development, and Library Technical Systems and Bibliographic Control program committees have established the Computer Files Task Force, which is charged with developing cooperative projects to improve access to computer files. The task force's initial efforts have been directed toward a need that emerged from the Stanford workshop discussions: to educate and train librarians in this area. A series of workshops, the first of which took place in May 1990, is designed to meet this need. As more libraries begin to acquire expertise in dealing with computer files, we expect to see additional committee-sponsored activities.

It is evident, however, that RLG alone cannot solve all of the problems identified. The consortium welcomes the opportunity to work with other organizations to bring computer files into the mainstream of research library operations.

—Constance C. Gould
Research Libraries Group, Inc.



Adding Computer Files to the Research Library: Issues in Collection Management and Development

By Margaret Johnson

Libraries have a long-standing familiarity with computer-based information technology. They began automating their operations in the 1960s and have been providing access to remote online bibliographic databases for nearly as long. But despite their technological sophistication, libraries have been slow to begin acquiring one of the most abundant products of information technology—the computer file.

Why is this so? Some believe it is caused by a failure to understand changing research needs and patterns. Practical considerations also undoubtedly contribute to libraries' reluctance to embrace computer files. The computer file, a dynamic medium that can contain textual, graphic, or numeric data, does not mesh neatly with existing library collections, structure, routine, and services.

Yet computer files have grown in their importance to research while research libraries and librarians have largely continued to ignore them. As a consequence, many researchers have become accustomed to circumventing the library and have devised their own methods for gaining access to such files through academic departments and direct contact with producers and data archives.

Academic libraries are already hard put to provide existing services; why should they take on this added responsibility? First and perhaps foremost, librarians are trained in organizing information and making it intellectually and physically accessible. They have the expertise to extend bibliographic control over all types of computer files. The library's strength is control of and access to information through standardized means—collocation through classification, subject, and names. Libraries are also expert at physically organizing and controlling huge inventories of material. They were quick to embrace computer technology as an effective and efficient means to achieve their twin missions of intellectual control and physical access. Providing these services in the area of computer files is a logical extension of existing responsibilities. Finally, researchers have come to expect libraries to direct them to relevant information regardless of format.

What will happen if libraries don't take the lead in computer file collection management and development? The library user will suffer and so will the library. Researchers will no longer be able to depend on the library as the central source of information about what is available and

Margaret Johnson is assistant director of the St. Paul campus libraries of the University of Minnesota.

how to obtain it. Without centralized coordination, expensive resources may be unnecessarily duplicated, even within the same institution, because both intellectual and physical access are limited. In addition, as libraries become less central to the research process, their position within the university may be degraded. Libraries may slip in institutional priority as planning and funding decisions are made. Above all, libraries will be distanced from an entire segment of the information universe—and one that is of growing importance.

How can libraries take on a leadership role? They should seek a central position in the selection, management, and delivery of electronic information in their institutions. Even if they elect not to assume a primary institutional role in managing and servicing computer files, libraries have much to contribute to this process. They can work in partnership with university administration, computing centers, and relevant research centers and academic departments. They need not be the sole proprietors of this territory but they can bring their particular expertise to the table as long- and short-range plans are developed. Indeed, expressions like "institutional information policy" and "information infrastructure" are meaningless if the library is not an active participant in policy decisions.

Once an academic library determines that it will take on collection management and development responsibilities for computer files, what steps must it take?

Understanding the Medium

To make informed choices about selecting and managing computer files, librarians need a basic understanding of the medium, its producers and distributors, and the possible uses of computer files. Like printed resources, computer files encompass a wide spectrum of information types, including numeric, bibliographic, textual, and graphic data. The term also covers software designed for instruction and data management purposes, and applications software. Thus computer files can be a resource for the delivery of traditional bibliographic information, a source of primary information, or a tool for organizing or manipulating electronic data.

Computer file formats range from magnetic tape to floppy and hard disks, optical discs, and CD-ROMs (compact disk–read-only memory). The kind of infrastructure of producers, distributors, and collectors that supports the dissemination of print materials does not exist for these media. The production of computer files is anything but straightforward. Producers include commercial firms, the government, professional associations, nonprofit organizations, and individual researchers. Commercial data files are compiled or created by publishers and marketed for purchase or lease; they are often termed *published computer files*. Privately produced files are generated by individual researchers, firms, or research teams and are not usually intended for public distribution; they are often referred to as *unpublished computer files*. Publicly produced files may come from nonprofit institutions, educational agencies, or consortia, and include large quantities of public-domain information from the federal government; depending on the completeness of the documentation and formality of the publication and distribution process, they are considered published or unpublished files.

Distributors include the producers themselves: individual researchers who have acquired such files; information services (e.g., the National Technical Information Service [NTIS]); information systems (e.g., state data centers); online vendors; organizations that archive data sets for public or member use (e.g., the Inter-university Consortium for Political

and Social Research [ICPSR]); data archives established by universities for their own faculty; and a few academic libraries (e.g., University of Florida Libraries).[1] The variety of ways in which computer files are distributed further complicates the process of gaining control of or even familiarity with what is available.

As researchers have begun to regard computer files as primary resources, institutional repositories other than libraries have evolved for these files. The need for such specially equipped repositories stems, in part, from the nature of the media. Magnetic tapes and hard disks remain the primary storage media for large numeric files, such as census tapes or survey research results. Files of this size require routine refreshing and updating of files, and organized collections of code books, introductory materials, and manuals. They also require trained operators and file managers to extract desired information from the available data.

An exception to the generally fragmented approach to computer files is the Inter-university Consortium for Political and Social Research, established in 1962 to manage the profusion of computerized information being collected in the social sciences. It serves as a central repository and provides dissemination services for machine-readable social science data (more than 25,000 files) and supporting documentation to researchers worldwide. ICPSR also offers training facilities in basic and advanced techniques of quantitative social analysis and additional resources that facilitate the use of advanced computer technology by social scientists. ICPSR is considered a model for other disciplines in identifying and controlling largely unpublished data sets, and has been effective because it is widely recognized among researchers in the social sciences. However, many ICPSR institutional memberships are held by academic departments and, again, bibliographic and physical access to ICPSR files at a campus are often not available beyond those precincts. The library and many researchers may be left out of the information loop entirely.

More commonly, large-scale data files are purchased and controlled by individual faculty, departments, or schools. Although some tapes may be housed in individual offices and studies, most are found in academic or administrative computing facilities—tape archives or tape libraries, data centers or data banks, and so on. Libraries have generally had neither the facilities and staff available to store, catalog, and maintain computer files, nor the inclination to do so. The result is a wealth of information resources within an institution, but limited information about what is available and thus limited access. Increasingly, libraries recognize that even if they do not intend to house all campus computer files themselves, they can perform a vital function in coordinating information about what is available and how to obtain access to it.

Identifying Appropriate Computer Files

Many research libraries have begun to acquire computer files under their own roofs. To do this effectively, the first order of business is to determine more precisely the scope of the universe of computer files, both published and unpublished. Identifying and locating computer files are a challenge. No single source or even a selected group of sources currently brings together reviews or lists of the files available. The routine reviewing infrastructure that exists for print materials has not yet been established for computer files. Few vendors, review journals, or inclusive directories are dedicated to computer files. Existing directories rarely list anything other than the mainstream, well-known, or commercially produced files. CD-ROMs, the type of computer file most frequently

marketed to libraries, are the most likely format to be reviewed in mainstream publications.

Keeping these limitations in mind, the librarian may find the following directories useful in identifying available files:

- *Computer Readable Databases: A Directory and Sourcebook*, edited by Kathleen Young Marcaccio. Detroit, MI: Gale Research, 1985. Lists 2,805 numeric and full text databases.
- *Data Base Directory*. (annual) White Plains, N.Y.: Knowledge Industry Publications, Inc. Lists approximately 2,650 online databases accessible in North America.
- *A Directory of Computerized Data Files: A Guide to U.S. Government Information in Machine-readable Form*. (annual) Springfield, VA: U.S. Department of Commerce, NTIS. Lists more than 2,500 numeric and textual data files from some fifty federal agencies.
- *Guide to Resources and Services: Inter-university Consortium for Political and Social Research*. (annual) Ann Arbor, MI: ICPSR. Lists over 25,000 files in approximately 1,600 titles held by the ICPSR.

These directories provide a starting point, but by no means describe the universe of computer files, especially when it comes to unpublished computer files. Unpublished files are especially difficult to identify and collect because they are produced outside the normal channels. Knowledge of them will be limited to those who created them and a few others with insider information, who may be unaware of their value to other scholars. Further, unpublished computer files often lack features associated with the usual publication process such as documentation and standardized bibliographic access. In addition, the medium is dynamic, allowing the file to be continually in development or under revision.

How can the librarian get inside information about computer files? The techniques are similar to those they have used to become familiar with any new field: reading the literature of the discipline for references to additional sources of information (in this case, computer files), participating in discipline-based professional associations and attending their meetings to keep abreast of new sources, networking with other librarians with similar responsibilities, and establishing regular contact with faculty to monitor their interests and projects. Eventually, not only will librarians work with their academic communities to identify the computer files that academics know they need, but they will also be in a position to bring new resources to scholars' attention—and either make them available or provide referrals to the individuals or organizations that provide access.

Several specific collaborative efforts could help librarians to define the universe of computer files. These include:

- Development of a formal network of librarians with responsibility for computer file selection and management or of libraries with significant computer file collections. This is already in the informal stage as word spreads through the profession that librarian X at institution Y is building a collection of computer files in subject Z. More formal networks will develop as librarians write articles and give presentations on their experiences, leading to wider recognition of individual expertise, and as discussion groups and committees focusing on computer files in libraries develop in the professional associations.
- Compilation of inclusive lists of available files into catalogs or directories of the *Books in Print* type, either in print or online format. This will

occur over time as the field matures and publishers and individuals realize a market exists.

- Creation of more consortia like ICPSR. This would provide subject-based forums for control of files and also retain that control within the academic community. Such a trend appears likely as the importance of computer files continues to increase across the disciplines.

Selecting Computer Files

Once librarians have begun to identify the computer files that are available, they must select from this universe. There are two schools of thought on how to manage computer file selection. One maintains that computer files are simply another among many media for the delivery of information and that the same standards and guidelines that direct all selection decisions can be applied. They believe that emphasizing the differences between computer files and more traditional formats will further separate computer files from the mainstream of library resources and services. The second school believes that the challenges associated with selection and management of computer files are unique or at least so unfamiliar that they require, if only in the short term, special attention.

Some libraries, among them those of Cornell University and the University of California at Berkeley, have taken the second position and developed computer file selection policies. In general, these policies reiterate the institutional collecting parameters already in place and go on to address the unique aspects of computer file selection and acquisition.

Few librarians propose ignoring generally accepted guidelines for selection decisions when considering computer files. Anthony Ferguson, in a paper on CD-ROM and collection development, presents eight criteria that Columbia University libraries apply when deciding what to add to the collection, regardless of format.[2] They are (1) relevance of material to program needs, (2) scope of treatment, (3) ability to fill existing gaps, (4) quality of scholarship, (5) currency of information, (6) accessibility of information, (7) language, and (8) cost. These criteria are usually embedded in the collection development statements or selection guidelines of most academic libraries. However, the criteria take on new, expanded meaning when the item under consideration is a computer file.

Relevance to program needs. Determining the relevance of computer files to program needs requires an expertise new to most selectors and bibliographers. He or she is expected to assess academic program needs in relation to the universe of computer files. This involves keeping abreast of published and unpublished computer files pertinent to his or her assigned subject area(s) and being aware of the computer files used and developed by the students and faculty in the appropriate academic programs. The computer file being selected should be within the scope of the collection and represent materials useful and important to a significant segment of the library's user community.

Scope of treatment. Scope of treatment refers to both information format and the depth and breadth of the material in the computer files. A library may decide that large numeric databases are more suitably acquired, housed, and managed by an academic department or the campus computer facility. However, the library may still apply selection criteria to these computer files in deciding whether the library should be responsible for providing bibliographic access to institutional files housed outside the library. A library may elect to exclude application software, except where the software is required to access and manage a file the library is acquiring.

It may exclude instruction software in the same way that some libraries exclude curriculum collections and textbooks.

Ability to fill gaps. Libraries have traditionally considered as an important criterion whether a particular item will fill existing gaps in the collection. Libraries will usually not purchase publications that are reprints of items owned unless an item needs replacement. Many computer files, particularly bibliographic files, duplicate print sources already in the collection. Yet these files do fill a gap in the collection in terms of improved patron accessibility and ease of use. Many computer files deliver information unavailable from other sources. Computer files should be an enhancement and enrichment of current collections and services.

Libraries have been cautious about canceling or withdrawing print resources that are duplicated by new computer files, in large part because of concerns about the stability of the various media. Certain computer file formats also present limitations in the number of simultaneous users or hours of service and accessibility, constraints not associated with multivolume print resources. In addition, some suppliers of CD-ROM products require return of all disks at the termination of the subscription or leasing agreement, leaving the library with no back files. Consequently, guidelines for acquiring duplicate resources must often be revised.

Quality of scholarship. The quality of the information and the product is not always easy to determine, particularly in the case of noncommercial and unpublished computer files. Ferguson speaks of evaluating scholarship, creativity, a work's lasting value, and the reputation of the author and publisher.[3] Bibliographers should expect the same excellence, comprehensiveness, and authoritativeness they require of materials in other formats. If the computer file is a commercial product from a recognized publisher or producer or has been reviewed, prepurchase evaluation is somewhat easier. For the time being, librarians must develop a certain amount of personal expertise as well as rely on colleagues and knowledgeable faculty to assess the quality of noncommercial files.

Quality of physical product. Determining the quality of the physical product presents new problems. Print materials usually receive a second review on receipt to ascertain if both intellectual and physical standards are met. It is not as easy to tell if a computer file has physical and data integrity or is even the file that was ordered. Some mechanism must be built into the review process to ensure that the computer file is, in fact, readable and that the data are accurate and reliable.

Currency of information. Until recently, currency of information was a greater problem than now. The production costs of computer files continue to drop and many commercial suppliers distribute updates in as timely a fashion as the print counterpart. CD-ROMs are among the more expensive formats and are less likely to be updated frequently than are mainframe tapes.

Mutability of information. A unique aspect of computer files is the ease with which they can be revised. Libraries that provide bibliographic access to computer files housed and managed outside the library must recognize that the file as originally described may not be the same file currently available. Although libraries may wish to avoid the custodial responsibility for archiving every version of local data files, it is important to recognize that subsequent revisions or generations of a computer file may obliterate earlier ones, leaving no record of the original data or history of the file's development.

Accessibility of information. One essential element in accessibility is adequate documentation. Undocumented data files are the computer-

ized equivalent of research notes and are of limited value. Documentation should be complete, comprehensible, and self-explanatory. It should guide the user to the information he or she wants. If a trained computer operator, programmer, or technician is required to extract data sets, the computer file must be accompanied by code books and the appropriate tag, field, and format information to manipulate the file. Libraries that acquire computer files must be prepared to provide the services necessary to make the information they contain accessible.

Many mainframe computer files and CD-ROMs can be acquired only on a subscription, lease, or contractual basis. These often have potentially restrictive clauses requiring a guarantee of limited access to ensure "fair use" for teaching, scholarly, and research purposes. Libraries may be required to monitor use and limit access to users who conform to certain criteria or to control downloading of data. This becomes troubling in public institutions that are funded to serve as state resources. Limits on accessibility also may present practical problems for the staff, who would be required to query all potential users of the restricted computer file.

The availability of computer files may be limited by patents, licenses, or other restrictions, although many private files are opened by their producers to other researchers on an ad hoc, informal basis. The proprietary nature of unpublished files is often ambiguous. When those files are created at an academic institution, questions arise concerning who owns the files and to whom outside that institution access should be provided.

Language. Language has not generally been an issue, but will become more important as more files become widely available. Certainly no library will acquire a full-text database in a language that is not within reach of a sufficient portion of its clientele. Numeric databases are universal, though problems may arise if the software and programming language required to access the data are nonstandard.

Cost. Cost factors associated with the selection of computer files are particularly far-reaching. Issues that require attention encompass the source of funds, how they will be allocated, and the continuing commitment involved. Both the start-up costs and continuing costs must be recognized. Computer files require, by definition, a computer to access the data, and the library must decide how to fund the initial hardware costs, including site preparation (workstations, wiring, telecommunication lines) and continuing costs (printer paper, ink cartridges, repair, and upgrade). Some computer files require separate access software or preprocessing. Libraries must recognize these allied costs and identify a source of funds to cover them.

Libraries have handled the cost of computer files in different ways. Several libraries have used grant monies or one-time institutional allocations to experiment with the medium. In nearly all cases, the result is so popular that patrons not only demand its continuation, but also request more extensive back files and additional files and services. Libraries then face the problem of continuing and increasing costs. Sometimes this provides a lever to approach institution administrators or academic departments for additional continuing funds. This can be politically complex: libraries often ask simultaneously for significant increases to keep up with the escalating costs of traditional print resources. Ultimately, all selection decisions must be evaluated in light of other potential acquisitions and other uses of the budget. Libraries must decide if this responsibility lies with the individual bibliographer, with a collection development committee, or at an administrative level.

Selecting a Variety of Formats

Another parameter should help to guide computer file selection: a strategy for selecting a mixture of computer file formats. Theoretically, a bibliographer should be able to apply existing collection development criteria when selecting among computer file formats. However, informed choices require a new knowledge base. For example, libraries may face choices between locally mounted mainframe databases, PC-based user workstations, mainframe gateway access to remote files, and librarian-mediated dial-up access to remote files—as well as between computer files and printed resources. Factors in such a strategy, in addition to those noted above, include availability of technology and technical feasibility and compatibility.

Preserving Computer Files

Preservation of machine-readable files raises new technical issues and cost implications for bibliographers' consideration. The selection and acquisition of most library materials carry an obligation to preserve the information in those materials in accessible form. However, no data permanence standards exist and the durability of many computer file formats is unknown, leading many libraries to retain paper copies (if available) as the preservation medium. A study by the National Archives and Records Administration recommends the use of paper or microfilm for archival purposes over optical and magnetic media.[4] It claims that the latter media lack the permanence that high-quality paper and microfilm offer. Magnetic disks and tape last only ten to twenty years, while the highest-quality paper and microfilm can last hundreds of years.

Guidelines for the preservation of magnetic tapes are fairly clear and involve proper conditions for storage, tape rotation, and systematic copying. These activities are, nevertheless, costly in terms of space, personnel, and dollars. Floppy disks are easier to manage, but are also more subject to destruction because of direct patron use. Libraries are beginning to develop policies and procedures to handle the back up, storage, circulation, and integrity verification of floppy disks. CD-ROMs cannot be easily backed up locally, but reputable commercial producers will often replace those that fail. In addition, there is a host of problems relating to file formats that become outmoded or are usable only on obsolete hardware. Reformatting is a potential solution but not necessarily the most appropriate. In the area of computer file preservation, the questions outnumber the answers.

Coordinating with Other Library Units

Including computer files in the collection development and management plan of libraries is complicated and intricate enough. The complexity of such plans is compounded by the fact that any single decision affects virtually all aspects of library services and operations. Although it is true that no library selection decision is made in isolation, the consequences of selecting computer files are more far-reaching and complex than those associated with more traditional resources.

The integration of computer files requires close cooperation between the different departments of the library. The following considerations should be borne in mind when making selection decisions.

Facilities planning. Selection decisions about the acquisition of a single print title seldom require the bibliographer to determine if sufficient shelving is available to house the item. In contrast, computer files and the hardware to access them do carry implications for adequate space, furniture, electrical and telecommunication connections, and physical security. Practical as well as fiscal considerations must be addressed.

Service and training. Bibliographers must work with public service units to identify how much and what type of librarian mediation

will be needed and how much is possible, given current staffing levels. The degree of assistance feasible should influence decisions regarding the type and content of files selected. A library may elect to exclude all computer files that require a trained technician or programmer to extract customized data sets.

Both library staff and users will require ongoing training. The type of training needed must be determined and responsibility for it allocated. To minimize needs, selection preference might be given to files with self-contained tutorials. Other decisions that should be jointly considered are the adequacy of available documentation and the possible need for locally developing additional documentation and "quick helps." Practical service considerations will normally reside with public service staff, but should be discussed as part of the selection process so that decisions can be modified and prereceipt planning can occur. These include decisions about scheduling CD-ROM workstations, responsibility for equipment (including troubleshooting and replacing printer paper and cartridges), security for hardware and software, and issues of contract-mandated controlled access.

Legal and contractual issues. The purchase agreement for many computer files is a formal contract with clauses that may restrict printing, downloading, storage, and transmission of data. Some libraries have implemented a policy that requires direct involvement of the library's fiscal officer and/or one or more upper-level administrators in the purchase process. This is to ensure that the library is prepared to accept responsibility for all legally binding restrictions of the purchase or leasing contract, or to allow the library to propose modification of the contract before signature by the fiscal officer or administrator. Copies of contracts may then be retained at a central location and university counsel consulted when necessary.

Acquisition. This process involves acquiring materials, selecting suppliers, assigning responsibility for contract review, and establishing procedures for quality checking on receipt. These are familiar operations, but they take on new twists when the materials are computer files. Bibliographers will need, at least initially, to work closely with the library's acquisitions unit to develop the necessary procedures and facilitate the processes.

Bibliographic control. Policies for bibliographic control should be explicitly formulated at the time the library begins to acquire computer files. Bibliographers should work with the cataloging unit to ensure that the level of bibliographic access is consistent with that for other materials in the collection. Ideally, this policy will state that all computer files acquired by the library for patrons, including those purchased with library materials budget funds, with grant money, or jointly with departments, and those received as gifts, are to be fully cataloged. Cataloging should be performed in accordance with all relevant national standards, including those for description, choice of and formulation of access points, recording of holdings, and representation of data in MARC (MAchine-Readable Cataloging) format. Exceptions may be made for files that cannot readily be mounted by catalogers for inspection; description of these materials may be based on documentation rather than examination of the files themselves. Libraries should provide for assignment of subject headings and classification numbers and any special handling that is necessary in the processing steps. It is important that bibliographers and catalogers recognize that computer files may present special problems and agree on how these will be handled.

Coordinating with Other Campus Units

Through computerization of their operations, services, and collections, libraries have become much more dependent on telecommunications, networking, and computing within the university. It is important that the library be actively involved in decision making regarding these facilities and campuswide services. Although this interaction should begin with library administration, it is important for those involved with collection development and management to be party to the process. Bibliographers should be familiar with the institutional resources (hardware, software, and computer files) if they are going to make decisions about the selection, management, and delivery of electronic information.

If computer files are located outside the library, bibliographers will need to be aware of their existence in order to avoid duplication. They should participate in discussions about placing these files under the bibliographic control of the library and about the selection criteria to apply in making such decisions. Regular and direct contact with academic departments is necessary for bibliographers to know what computer files are available and important to their subject areas.

Training Bibliographers

As noted above, the only effective way to begin the process of integrating computer files into collection development and management activities is through staff education. Several participants in the Research Libraries Group's Machine-Readable Data File Project began their activities with a series of in-house introductory sessions on computer files for library staff.

Although it is possible to argue about how much technical expertise is necessary to make informed selection decisions, it is clear that bibliographers need a basic understanding of computer files to begin giving the medium serious consideration.

Education begins with consciousness raising about the value and importance of computer files as research and resource tools. It should present information about practical and at least elementary technical considerations. Other points addressed in this paper, including identification, selection, and coordination issues, should be part of the formal training of bibliographers as well. Bibliographers need as much information and understanding as possible to make the best decisions for their institution.

Several avenues are open for training. Libraries can plan a series of local sessions, drawing on in-house expertise or bringing in outside authorities. Regional and national workshops can be organized to address shared deficiencies. Library schools must include all aspects of computer files as both resource and research tools in their curriculum, as a matter of course.

Preparing for the Future

Libraries, in their central role as providers and organizers of information, cannot afford to ignore computer files or to approach them in a piecemeal fashion. To do either places the library at risk of becoming less valuable to and less supportive of the academic community. Libraries will no longer be regarded as the focal point for information as more of it slips outside their purview. Researchers will no longer think of the library as the first resort because they will be unable to depend on its catalogs, collections, and directional tools to provide access to the universe of information resources. The library's significance to the administration will diminish as well.

Librarians do not have to allow this to happen. They have skills and expertise in the organization and delivery of information. They have worked with computer technology for years. In short, they are

admirably positioned to develop and manage computer file collections and services.

Librarians must chart a bold and visible course of action in dealing with one of the primary challenges to the continued value and service of research libraries. ■

References

1. Lyn MacCorkle, "Data Archives, Data Sets and Libraries," *Library Hi Tech Bibliography* 3 (1988): 45-56.

2. Anthony W. Ferguson, "Assessing the Collection Development Need for CD-ROM Products," *Library Acquisitions: Practice & Theory* 12 (1988): 325-332.

3. Ibid., p. 327.

4. *Preservation of Historical Records—The National Archives & Records Administration Study of Preservation Media* (Washington, D.C.: National Academy of Sciences Press, 1986).

Additional Reading

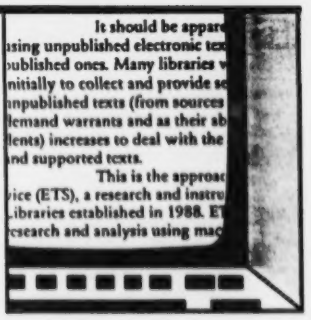
Michael K. Buckland, "Foundations of Academic Librarianship," *College & Research Libraries* 50 (July 1989): 389-396.

Kathleen M. Heim, "Social Science Information Needs for Numeric Data: The Evolution of the International Data Archive Infrastructure," *Collection Management* 9 (Spring 1987): 1-53.

Ray Jones and Thomas Kinney, "Government Information in Machine-Readable Data Files: Implications for Libraries and Librarians," *Government Publications Review* 15 (1988): 25-32.

Bliss Beckman Siman, "Machine-readable Data Files in the Library," *Technicalities* 6 (June 1986): 9-10,+.

David Tyckoson, "The Cost/Usage/Accessibility Formula: Selecting an Information Format," *Technicalities* 8 (June 1988): 7-9.


It should be appare
using unpublished electronic tex
published ones. Many libraries v
initially to collect and provide se
unpublished texts (from sources
demand warrants and as their ab
llents) increases to deal with the
nd supported texts.
This is the approac
vice (ETS), a research and instru
Libraries established in 1988. ET
research and analysis using mac

Machine-Readable Texts in the Academic Library: The Electronic Text Service at Columbia University

By Anita Lowry

Recently members of the HUMANIST online discussion group engaged in a spirited discussion. Argument centered around the significance of a recent publication of Shakespeare's complete works on CD-ROM and whether libraries would purchase such a product. In the opinion of one member of the group, "the crux of the matter is the function of the library in research. Is the library a place where one can actually *do* research, or is it only a place where one can look up the research that others have done?" He notes that libraries primarily buy indexes in electronic form, rather than text files, and that "these products permit the scholar to find out very quickly what hypothesis other scholars have pursued or are pursuing. Products like the electronic Shakespeare, on the other hand, might actually be used to *do* research." He points out that they enable scholars to answer questions like, "Do the words *right* and *blood* ever occur in the same speech in any of Shakespeare's history plays?"[1]

This correspondent states the case for machine-readable primary sources in the library, treating it as the "laboratory of the humanist."[2] His argument rests on the assumption that machine-readable texts are valuable research tools for scholars and students in the humanities. Until recently this assumption might not have been shared widely by librarians, but it is gaining increasing acceptance in the scholarly world as electronic text files become more accessible, enabling scholars to discover their utility and put them to a variety of uses.

Electronic texts and text databases are, on the one hand, reference tools that can be used by librarians and patrons to look things up in ways that would be impossible without computer assistance. They can also be sophisticated analytical tools for an individual scholar's unique lines of inquiry, in conjunction with software whose capabilities are not limited to information retrieval or concording functions.

Accepting the premise that machine-readable texts are valuable for research and belong in the library, or should be accessible through the library, raises a host of questions and challenges for librarians. Many of these questions relate to the provision of access, reference, and instructional services for electronic primary sources. While a sizable body of literature examines service issues for machine-readable numeric data files in the sciences and social sciences, none yet exists for primary source textual files in the humanities.

Although machine-readable texts in the humanities differ significantly from numeric data files in content, nature of use and users, software for access and manipulation, and so forth, there are similarities

Anita Lowry is deputy head of the Butler Reference Department and director of the Electronic Text Service at the Columbia University Libraries.

between textual and numeric files in the types of library service issues they raise. Control of and access to the machine-readable item and to the equipment required to use it, difficult enough in themselves, are complicated by security, preservation, and copyright issues. Use of textual or numeric data for either reference or research purposes, often entails extensive instruction in and assistance with the content and structure of the file, the operation of hardware and complex software, and integration of information and methods from a variety of sources. For libraries to provide services for machine-readable files, the jobs of both professional and paraprofessional public services staff will need to be redefined, and these staffers will need to learn many more things than were dreamed of in "Reference 101."

How librarians respond to these challenges will define the future role of the library in humanistic research and influence the way such research is conducted.

Published vs. Unpublished Electronic Texts

In considering electronic texts it is helpful to make a distinction between published and unpublished texts. This distinction has nothing to do with whether the source text represented in electronic format was originally a published book or an unpublished manuscript. Rather, it defines a published electronic text as one whose goals and processes of production and distribution parallel the goals and processes that we associate with formal publication. That is, the electronic text is intended for distribution or sale to an audience beyond its creator. The text and the analytical software that may be supplied with it are subject to editorial control to ensure completeness, accuracy, and usability. The text is documented regarding source, modifications made, encoding and mark-up or tagging schemes used in the text, and so on. Instructional manuals are provided. Technical support for the text and accompanying software is available from the publisher.

During the last few years, pioneering nonprofit publishers of electronic texts, like the *Thesaurus Linguae Graecae* Project at the University of California at Irvine,[3] the ARTFL Project at the University of Chicago,[4] and the Center for Computer Analysis of Texts at the University of Pennsylvania,[5] have been joined by companies like the Electronic Text Corporation, Oxford University Press, the InteLex Corporation, and Niemeyer Verlag in preparing and distributing electronic texts.

In contrast, an unpublished electronic text is created by an individual or a research project without being intended for distribution beyond its initial users. Therefore, while such a file may serve its creator's purpose admirably, it may not have the documentation, explicit quality control, and support that would make it easy for others to use. And unless it has been deposited with a major text archive that has a formal distribution program, it will not be readily available for purchase.

Thousands of unpublished electronic texts exist in the hands of scholars, institutions, and book publishing companies (i.e., their computerized photocomposition tapes). Some of these unpublished electronic texts have been deposited in electronic text archives, like the Oxford Text Archive (OTA) at Oxford University. Although the OTA will, for a fee, distribute copies on magnetic tape or diskette of texts for which it holds distribution rights, it is not a publisher in the sense defined above, and many OTA texts lack adequate bibliographic identification for the edition of the source text used, have no explanation of the encoding and tagging schemes employed, and vary greatly in terms of accuracy.[6]

Whether published or unpublished, electronic texts suffer from inadequate bibliographic control. There is no *Books in Print* or *Cumu-*

lative Book Index, no *National Union Catalog* for electronic texts or analytical software programs, making it extremely difficult for librarians to find out what texts exist, in what format, where they are located, and whether they are available for purchase. Among the many partial sources of information about electronic texts are the Rutgers Inventory of Machine-Readable Texts in the Humanities (accessible on RLIN®, the Research Libraries Information Network); publications (and meetings) in the field of humanities computing, like the *Humanities Computing Yearbook*; lists of text archive holdings, like the *Oxford Text Archive Short-Title List*; and flyers from publishers, like the Electronic Text Corporation. Although never a simple, straightforward task, it is, of course, easier to gather information about published electronic texts than unpublished ones.

It should be apparent from this discussion that acquiring and using unpublished electronic texts is generally more problematic than for published ones. Many libraries will find it easier and more appropriate initially to collect and provide services for the latter, gradually adding unpublished texts (from sources on their own campuses and elsewhere) as demand warrants and as their ability (and that of their faculty and students) increases to deal with the more idiosyncratic, less well documented and supported texts.

This is the approach we are taking at the Electronic Text Service (ETS), a research and instructional facility of the Columbia University Libraries established in 1988. ETS activities and resources focus on research and analysis using machine-readable texts, the creation of bibliographic databases, and computer-based literature searching. At present, the equipment and collections of the Electronic Text Service include microcomputer hardware and software, machine-readable texts, database and hypermedia research tools, printed reference materials and documentation, and a small classroom area. The ETS enables staff to collect, examine, and evaluate software and text files, consult with students and faculty, and teach courses and workshops on new information technologies. It is designed to help Columbia faculty and students incorporate computer-based textual and bibliographic information into their research, study, and teaching.

During the first two years of the Electronic Text Service, we faced innumerable questions regarding infrastructure, collection development and management, access policies and procedures, bibliographic control, staffing and staff training, reference and referral services, instruction, and, of course, funding. None of the questions had easy answers, some have not yet been answered at all, others have answers that are evolving over time as we gain more experience and knowledge. What are some of the key questions relating to library services for electronic texts?

Reference, Research, and Instructional Services

Reference librarians have traditionally divided reference queries into categories like "directional," "ready reference," and "real reference questions" that express differences in the nature and difficulty of the question and the type of collection or amount of professional expertise required to answer it. Data archivists/librarians have employed a similar progression of categories for the types and levels of service associated with machine-readable numeric data files, from the least extensive ("passive" or "referral") to the most extensive ("maximum" or "expert"),[7] as an aid both to defining collection development and service goals and to determining what kind of staff is needed to provide a given level of service.[8]

What are the comparable categories of service for electronic texts, and what kinds of activities and skills do they encompass? What do

humanities scholars want to do with electronic texts, and what kinds of help do they need? What are the relevant characteristics of electronic texts and text analysis software that determine the skills needed to deal with them? What can existing, off-the-shelf texts and software do, and what requires custom programming and data manipulation? How will librarians' increasing involvement with electronic primary sources (as opposed to bibliographic and tertiary sources) and with advanced instructional activities change their role in the research process? What levels of service and staff expertise will (can) the library support?

Beyond these questions, equally difficult ones remain. Who does what? How do we redefine jobs and develop the requisite staff expertise? How can we work with other professional staff on campus, in the computing center, or in special humanities computing support services, to share expertise? How do we do all this at a time when libraries' professional staffs are not growing and when demands on staff time are increasing on all fronts?

These are questions waiting, begging, to be answered. A major goal of the Electronic Text Service is to provide the conceptual and physical framework for finding and testing the answers and their profound implications for libraries. For how can we understand the full range of scholarly uses for electronic texts and identify which uses are most prevalent and important if our faculty and students—and we—do not have access to them?

In the ETS we hope to test several basic assumptions:

- That a major scholarly application for electronic texts will be in looking for evidence, identifying ideas, themes, images, motifs, linguistic features, etc., as expressed in a text or body of texts, and seeking patterns in those features.
- That librarians' skills in the principles and techniques of computer-based information retrieval, especially in full-text databases, uniquely prepare them to master complex searching software and retrieval problems.
- That since many academic librarians have subject backgrounds in the humanities and knowledge of foreign languages, they will be able to understand humanists' information needs and deal with electronic textual data.
- That as humanists are exposed to the power of computer-based information retrieval through using CD-ROMs and online library catalogs, as they draw conclusions about the potential for manipulating texts from their experience with word processing, as more electronic texts become available, and as scholars are led to them by colleagues, scholarly societies, and librarians, then students and scholars now generally ignorant of the existence of such texts and their potential uses will begin to regard them as indispensable.
- And, finally, that humanities scholars and students are going to need a lot of assistance and instruction if they are to take full advantage of electronic text resources.

Given this last premise in particular, we expect that all reference librarians in the Butler Reference Department (of the Columbia University Libraries) will know the ETS collections and how to use the most general resources. But we have also divided responsibility for in-depth knowledge of text databases and analysis software among the reference

staff members in accordance with their language and subject expertise. Responsibility for developing instructional activities and materials for various resources in the ETS is similarly assigned. A comprehensive training program for librarians is planned, but staff members have already taken the initiative in learning about the particular resources for which they are responsible.

For users, the cornerstone of our instructional program is a 1.5-credit course taught each semester for students in the humanities and history departments of the Graduate School of Arts and Sciences. "Research in the Humanities: a Practicum on Resources and Methods" introduces graduate students in the humanities disciplines to "resources in print and electronic formats that are fundamental to advanced humanities research. Topics covered by the course include major reference tools, including computerized online and CD-ROM databases; microcomputer databases for managing personal notes and bibliographies; scholarly communication and publishing; machine-readable texts, textual analysis, and critical editing."[9] Two classes are devoted to electronic texts: the first one introduces these new research tools, and the second class divides students into small groups by discipline for more concentrated instruction on a particular electronic textual resource.

These are key building blocks for the services of the future, and they complement the more concrete foundations we are erecting for the physical infrastructure of the Electronic Text Service.

Infrastructure

What kind of space, facilities, and equipment are required to provide access to electronic texts?

When planning for the necessary infrastructure, we must first consider the characteristics of the collection itself: the texts, software, and supporting materials. Although electronic texts originated nearly 40 years ago in the use of computers to compile concordances,[10] their coming of age dates from the advent of the microcomputer.[11] At present, most electronic text publishing is microcomputer-based: CD-ROM and magnetic diskette are the preferred distribution media, and most specialized software development in this field is currently being done for microcomputers. (A number of good, relatively easy to use, textual analysis programs already exist for IBM-compatible microcomputers.) The ARTFL and Dante text databases are available online, via SprintNet and the Internet, and are thus accessible to users through microcomputers. In addition, important hypertextual and hypermedia research tools, like the Perseus Project at Harvard,[12] are being developed for distribution on CD-ROM and laser disk for use with Macintosh microcomputers. Accordingly, the Electronic Text Service houses both IBM-compatible and Macintosh microcomputers, CD-ROM and laser disk players, a Bernoulli Box (a hard disk drive with removable hard disks to accommodate an unlimited number of text databases), and telecommunication connections to the outside world.

Configuration of space and equipment should be determined by the use to which the electronic resources will be put and by the services they will require. Though the ETS has not yet fully achieved this ideal, it is highly desirable for electronic text facilities to be located close to reference staff and to relevant reference materials (like foreign language dictionaries), as well as to the libraries' research collections of printed texts. This facilitates reference use, staff support, and both casual and intensive use of the electronic resources in conjunction with traditional library sources.

Because instructional services for electronic texts are paramount, space and facilities must be provided for both group and indi-

vidual instruction. In addition, adequate, semiprivate work space is needed for students and faculty members to use particular resources for their study or research. Within the current small quarters of the Electronic Text Service, librarians can teach a class of approximately 20 people (using an LCD projection panel to project the microcomputer screen images onto the wall), they can consult with individuals at any of the three workstations, and individuals can work at a particular workstation—but because all facilities are housed in one room with no dividers, only one "noisy" activity can be going on at a time.

What about remote access to texts via networks? Librarians' experiences with CD-ROMs and with the online catalog have sensitized them to the potential value of multiuser, remote access to electronic information. But when we consider providing access to electronic texts over LANs or other campus networks, we raise a host of new questions and problems. In particular, the licensing agreements for most published electronic texts and analytical software specifically prohibit mounting them on a network or multiuser system; special licenses for multiuser access, when available, are significantly more expensive.[13] Secondly, while most mainframe-based systems—like a library online public access catalog—can accommodate a basic full-text information retrieval software program, they may not be able to deliver access to specialized textual analysis programs of the kind currently being developed for microcomputers.[14] Thirdly, remote users are particularly difficult to reach with instruction and assistance. These and other problems may not be insurmountable. But, at least in the near term, the Electronic Text Service is concentrating its resources and efforts on the variety of texts and software available for microcomputer-based, single-user access.

Finally, no discussion of infrastructure would be complete without some consideration of security issues—both security for physical facilities and for data. With CD-ROMs, libraries have been forced into a delicate balancing act between protecting the equipment, data, and software from theft, misuse, and damage on the one hand, and providing optimal, user-friendly access on the other; an over-emphasis on one may unduly compromise the other. The interplay between security measures and access procedures is equally important for electronic text collections; librarians will look to their experience with CD-ROMs and other electronic media and equipment and to the experience of data archives and computer centers for guidance.

Access Policies and Procedures

Given the concern about security and users' needs for instruction and assistance, can a library provide the equivalent of open stack access to electronic text facilities some or all of the hours the library is open? Can its staff give extensive reference and instructional service on demand or only by appointment? Will users be charged a fee for use, for printing, for authorized downloading or copying? Can a library afford to offer access and service to people outside its primary clientele groups? Are electronic texts and software to be circulating or noncirculating? If circulating, how to manage it? How about interlibrary loan? Will users be able to dial up remote online electronic text databases like ARTFL from anywhere on campus or only from the library? Will there be a central gateway to such databases as part of the library's online information system, and will there be a charge for their use? If texts are mounted on a campus network, are they accessible only to those with local system IDs or to anyone on the Internet?

Access policies and procedures for library materials often vary among libraries and are determined by such things as the nature of the collection, size and characteristics of the user population, size and composition of the public services staff, physical layout and facilities, even the geographic location of the library. How a library answers the questions just posed will depend on many variables. Perhaps the most important are (1) the licensing and access agreements imposed by the publisher or distributor; (2) the need to protect the collection from theft or damage; and (3) the number of reference librarians and other staff available to provide service.

At present, the access policies of the Electronic Text Service are fairly restrictive. Long-term objectives are more liberal—for example, networked as well as walk-in access and on-demand services for at least some of the ETS resources at least some of the time—as staffing and facilities permit. We will try to realize the ideal of making access to machine-readable texts as direct, convenient, and successful as possible with a minimum of physical, procedural, and economic barriers, our goal being to integrate electronic primary sources into the collections and services of the library, and thus into the research and teaching of the university community.

The Long-term Goal

Integrating electronic texts fully into the library and into the scholarly enterprise is an ambitious and long-term goal. Achieving it will require that we address many unanswered questions and attain many intermediate goals and objectives. At the same time, new developments in technology may change the way some of those questions and goals are framed; to a certain extent, we are aiming at a moving target. But if scholars are to see the library as a place where they do research, we must continue to adjust our sights and services until we succeed in creating a "humanist's laboratory" for the future. ■

Notes

1. Brian Whittaker, "Oxford Electronic Shakespeare, etc., cont.," electronic message on *HUMANIST Mailing List*, Vol. 2, No. 740 (18 March 1989) (electronic Listserv discussion group *HUMANIST@UTORONTO*, moderated by Willard McCarty, *MCCARTY@VM.EPAS.UTORONTO.CA*).

2. The metaphor of the library as the humanist's laboratory is a commonplace. But this metaphor has rarely received more extended explication than by Bernhard Fabian in his article "Libraries and Humanistic Scholarship" (*Journal of Librarianship* 8 [April 1986]: 79-92). Although Fabian does not mention machine-readable texts, it is difficult not to envision the potentially immense contributions of electronic texts to the research process he describes, both by providing direct access to large bodies of diverse texts and by facilitating the humanist's identification and exploration of relationships, patterns, and distinctions among them.

3. "The Thesaurus Linguae Graecae (TLG) is an electronic repository of Greek literature from Homer, in the eighth century BC, to the closing of the Academy in the sixth century AD, and expansion to AD 1453 is now underway. The TLG consists of a data bank of approximately 3,000 authors, representing more than 8,000 individual works and about 61 million words of Greek text. . . . The Committee on the Thesaurus Linguae Graecae of the American Philological Association has approved the editions for the TLG texts on the basis of scholarly quality, date, and availa-

bility. As better critical editions are issued, they are encoded and substituted for their older counterparts in the data bank." (*The Humanities Computing Yearbook, 1988* [Oxford: Clarendon Press, 1988], 183-4.)

4. While the ARTFL (American and French Research on the Treasury of the French Language) Project does not "sell" electronic texts, its online database—currently consisting of approximately 2,000 French literary, historical, philosophical, and scientific works published between the seventeenth and twentieth centuries—is produced and accessible according to the criteria for publication established in the discussion above. The Dartmouth Dante Project (producing an online database of the texts of *La Divina Commedia* and major commentaries) at Dartmouth College and the Global Jewish Database project (creating an online database containing the full text of the worldwide "Responsa literature," the Hebrew Bible, the Babylonian Talmud, the Midrash literature, Maimonides' Code, etc.) at Bar-Ilan University in Israel are similar publishers of online electronic texts.

5. In conjunction with the Packard Humanities Institute, the Center for Computer Analysis of Texts (CCAT) produces and distributes the *PHI/CCAT* CD-ROMs containing classical Latin texts and texts relating to ancient and Biblical studies.

6. See Judith K. Proud, "The Oxford Text Archive," *British Library R&D Report*, no. 5985 (February 1989): esp. 12-19.

7. Cf. Laine G.M. Ruus, "User Services in a Data Library," *IASSIST Newsletter* 4.2 (1980): 29-33; Jim Jacobs, "Some Levels of Library Data Services," notes distributed at IASSIST Conference, Washington, D.C., May 29, 1989; Edward P. Bartkus, "Use of Numeric Databases in Reference and Information Services," *Drexel Library Quarterly* 18 (Summer/Fall 1982): 205-219; Ray Jones and Thomas Kinney, "Government Information in Machine-Readable Data Files: Implications for Libraries and Librarians," *Government Publications Review* 15 (1988): 22-35; Katherine Chiang et al., "Computer Files in the Research Library: Issues of Collection and Access," in *Managing a New Library Resource: Results of the RLG Machine-Readable Data File Project in Six Major Libraries* (Mountain View, CA: The Research Libraries Group, Inc., 1989), 289-423, esp. 323-337.

8. The Mann Library at Cornell University, one of the boldest and most articulate advocates of "mainstreaming" electronic data files, has taken a systematic and well-thought-out approach to defining the tasks and skills required to provide extensive support for numeric data files, translating these into new and redefined public services job descriptions, and hiring librarians and technical staff to fill these positions.

9. Anita Lowry, "Beyond BI: Information Literacy in the Electronic Age," *Research Strategies* 8.1 (Winter 1990): 23.

10. In the 1950s, Father Roberto Busa used a computer to produce the massive *Index Thomisticus*, "a sophisticated computer-produced linguistic analysis of 118 writings of St. Thomas Aquinas and of 61 other works associated with the *corpus thomisticum*, documenting the vocabulary and usage of 179 Latin works from the 9th to the 16th century." (Eugene P. Sheehy, *Guide to Reference Books*, 10th ed. [Chicago: American Library Association, 1986], entry no. BA129.) See also Dolores Burton, "Automated Concordances and Word Indexes," *Computers and the Humanities* 15 (1981): 1-4, 83-100, 139-154; 16 (1982): 195-218.

11. For example, although the *Thesaurus Linguae Graecae* texts were available for many years on magnetic tape, it was not until the TLG was released on CD-ROM (along with a comprehensive microcomputer-based workstation and software) that it became sufficiently affordable, accessible, and easy to use to be integrated into the scholarly life of the discipline. Now the TLG is located in classics and religion departments, in libraries, and even in the offices and homes of individual scholars throughout the U.S.

12. "The Perseus Project is an ongoing effort to amass a large hypermedia database of materials pertaining to classical Greece. Its object is to build an electronic library with a broad spectrum of information useful for many different purposes, primarily in research and teaching . . . [Perseus will contain] Greek texts and translations, the Intermediate Liddell-Scott Lexicon, colour video images of archeological [sites] and objects, topographical maps of Greece annotated with toponyms, . . . an historical overview and a classical encyclopedia. . . ." (*Tools for Humanists, 1989* [Toronto: Centre for Computing in the Humanities, University of Toronto, 1989], 85-6.)

13. For example, the *Thesaurus Linguae Graecae* CD-ROM license explicitly forbids putting it on a file server or accessing it on a network; InteLex's license prohibits using its search software on a network and limits use of its texts to "not more than two (2) single computers at a single location"; both the Electronic Text Corporation and Oxford University Press licenses for software and texts limit use to a single computer.

14. This is not to say that there are no sophisticated analytical tools for text available on mainframe computers. But they tend to be utility-like rather than fully developed end-user programs; they are often batch-oriented programs; and those specialized programs designed explicitly for humanities textual applications are supported by relatively few campus mainframe computing facilities.

Bibliographic Description of Computer Files: Lessons Learned from a Case Study

By Lynn Marko

In recent years, librarians have become increasingly aware of the need to extend access to materials beyond books and serials. In particular, they are beginning to look toward providing access to data in electronic formats. But although the library profession has developed effective and efficient methods for handling print materials, it is less well equipped to handle materials in machine-readable form. Librarians are discovering that the realm of electronic information is not easily accessible, and that there are few road maps to this *terra incognita*.

How does a research library begin the journey into this new realm? For the University of Michigan, an opportunity to do so arose in 1987 when the Research Libraries Group, Inc., proposed a cooperative project to catalog a large collection of machine-readable social science research data.

The goal of the project was to make information about the data archives of the Inter-university Consortium for Political and Social Research (ICPSR) available online to the research community through the Research Libraries Information Network (RLIN®). A central repository and dissemination service for machine-readable social science data, ICPSR collects, processes, and distributes machine-readable data on social phenomena in more than 130 countries. Sources of data include large organizations such as the United Nations and the U.S. Bureau of the Census as well as research institutes and individual researchers.

The Project

To provide an access resource for the user community, ICPSR has prepared and maintained the ICPSR *Guide to Resources and Services*—a printed guide used as a means of controlling and providing access to ICPSR holdings. The guide was produced from a SPIRES* database containing information supplied by the original creator of the data and was designed to meet the needs of the specific community of the consortium members. Both to control and to transfer the members' own local files as they became larger, ICPSR wanted to develop a more generalized access system than that which was then available to the social science community.

Over the course of six months, 1,300 SPIRES records for data sets from the ICPSR *Guide to Resources and Services* were entered by tape transfer into the RLIN® database and converted into the USMARC record format in this process. (Although the ICPSR staff were aware of MARC,

*SPIRES is a database management system used in a number of universities and colleges.

Lynn Marko is the head of monograph cataloging at the Hatcher Graduate Library of the University of Michigan.

they had felt it was necessary to develop a rule-based cataloging system unique to the social science data community.) The University of Michigan Library cataloging staff reviewed each record in RLIN® for cataloging appropriateness, normalized all field tags and indicators to make each record conform to nationally accepted cataloging standards, provided name headings in AACR2 form, and supplied Library of Congress subject headings. The Research Libraries Group provided the technical support to transfer SPIRES data to MARC format via tape transfer.

In constructing the note fields in the reformatted records, we (the University of Michigan Library cataloging staff) relied entirely on the summary statement in the original records, which varied in fullness from cryptic to exhaustive. We moved data into the appropriate special note fields, trying not to eliminate any of the original information but in some cases revising it for the sake of clarity. We added information when it could be found, for instance, citing the title of the computer file code book if it differed from the study title. And, in some cases, we had to guess at information we added, such as the publication or release date. We retained notes citing the ICPSR holding archive and the ICPSR data class and added a standard note citing availability of data and documentation from ICPSR, including full address and telephone number. The cataloging staff considered it worth the effort to sort the available information into special MARC fields because in the RLIN® system virtually every field is accessible via free text searching.

We supplied the standard access points for bibliographic records, including the names of principal investigators (up to three) and corporate body principal investigators, adding in many cases the names of individuals and agencies involved in the collection of the original data or preparation of the machine-readable form of the data. We made every effort to supply these access points in established forms so that the headings would be compatible with headings used in records for books, serials, and other library materials. Library of Congress subject headings were assigned to each record in addition to the ICPSR keyword subject terms already present in the record. The goal was to achieve compatibility with records for other types of library materials.

Challenges Encountered

The first challenge was to coordinate access to materials not actually available for review. When a book is cataloged, the cataloger takes the information directly from the printed material. A computer tape or diskette lacks easily verifiable information about what is written on it. At best, it is accompanied by a code book or other documentation that provides information about what is on the tape, where on the tape the information is located, how it is coded, and so on. However, code books are working documents put together by the compiler of the file, and can be as informal as a handwritten notebook—and some computer files are not accompanied by even minimal documentation. If the library has the hardware and the expertise required to mount a tape, it can examine the file and attempt to verify whether it contains the information described in the code book. But in our case, the sole source of information was the ICPSR guide record as loaded into the RLIN® database with the SPIRES field names converted into MARC numeric tags by an RLIN® conversion program. Because of time and other practical constraints, it was impossible for us to run any of the computer files or read code books to obtain more information.

Another challenge was the absence of established practices to rely on for the bibliographic description of large computer files. In fact,

the descriptive cataloging of machine-readable data files was in such an unsettled state that *Anglo-American Cataloging Rules*, second edition, Chapter 9, was completely revised midway through the project, requiring several major changes in order and content of information required. In spite of the publication of the draft revision of AACR2, Chapter 9, Michigan retained its established record design so as to be consistent throughout the project, but one change we were pleased to adopt was the new generic term *computer file* in place of the longer term, *machine-readable data file*, in use earlier.

Assignment of subject headings also presented difficulties. The keyword terms assigned by social scientists or ICPSR staff and included in the ICPSR guide records were very specific and did not always point to an appropriate Library of Congress subject-heading term. In assigning subject headings, we constantly asked ourselves how broad or how specific they should be and what number of subject headings was reasonable or practical.

In short, the project was much more complicated than anticipated at the outset—but the rewards were significant. The experience enabled us to free ourselves from the book model of bibliographic description and to become more confident in describing computer files. We now create collection-level records for major CD-ROM and online products available at the University of Michigan. These include the ARTFL (American and French Research on the Treasury of the French Language) database, the *Oxford English Dictionary*, Supermap, and software packages.

In so doing, we are continually presented with new challenges. For example, the ARTFL database record created for RLIN® has 27 subject headings, and the *Thesaurus Linguae Graecae* record has 70. We are pushing the limits of MARC, adapting it, and trying to take advantage of its flexibility rather than chafing at its constraints. This process has helped us to shape an institutional plan for computer files at Michigan and has led to Michigan being a beta test site for Multiple Database Access System applications of NOTIS. We were one of the first libraries to mount non-bibliographic index databases, such as the Wilson indexes and PsycINFO, on our network as part of the library online services to the campus.

Issues

In the course of this project, we encountered but did not resolve issues that are applicable to the bibliographic description of all computer files.

Cataloging without a title page. The primary mission of catalogers has been to provide generalizable access to traditional materials. By their very nature, computer files are not traditional. A computer tape, diskette, or compact disk does not present the same information as a title page. Without the benefit of a title page or something analogous to it, looking for descriptive information is time-comsuming and sometimes fruitless. Staff at Michigan questioned whether some data elements important for books are relevant to computer files and worth the time spent ascertaining them.

Dealing with primary data. Catalogers, by and large, are more accustomed to dealing with interpretive data than with primary data. Primary data has become the specialty of archivists, who have developed techniques to collocate large bodies of primary material in ephemeral form. They provide collection-level access for the scholar, who can then determine where to begin his or her research.

Providing subject access. Assigning subject headings is difficult because many terms in use are not broad enough to describe large amounts

of primary data. However, index terms, usually generated without benefit of a thesaurus, can be too specific and too plentiful. As records become more numerous, do we want to load into our systems records with many uncontrolled terms? When records are very long, how should they be edited? What information is important to a specific community of users and what information is valuable to general users?

Dealing with a mutable medium. Traditional bibliographic description is container-based, an orientation that creates difficulties in describing computer files, which are mutable with regard to both format and content. Data can be easily transferred from one electronic format to another. The content can also be easily altered, but no distinction is made between major and minor revisions. The concept of edition for a computer file is not well defined. How and when do we tell the user that the content has been changed? How do we know?

Revising records. The records themselves may need to be revised. Updating bibliographic records for materials as dynamic as computer files, and incorporating them into the ever-changing environment of local and national bibliographic databases, is difficult. For example, when we created the ICPSR records in 1988, a number were actually revised versions of records issued earlier by ICPSR. The choice had to be made of either overlaying the already upgraded RLIN® record with a revised SPIRES record from ICPSR or missing the benefit of content revisions that ICPSR had in these SPIRES records. RLG chose to overlay the newer ICPSR record on the upgraded RLIN® records. As a result, all the catalogers' additions, changes, and enhancements were lost and had to be made again. We subsequently reached an arrangement with ICPSR to send us printouts of ICPSR's SPIRES records with revisions noted. We then make the changes directly online in RLIN®.

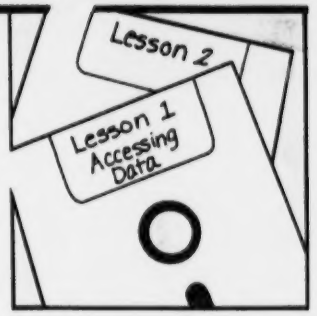
Providing access but not service. In many cases, the library neither owns nor provides service for files it catalogs. At Michigan, the ICPSR institutional membership is held by an academic department rather than the library. In creating bibliographic records and providing access to this information through the local online catalog, the library is a partner with the academic department and ICPSR to bring these files to the attention of users.

Restricting user access. User access restrictions determined by facilities, local support, and general campus priorities need to be explored more fully as the materials described become better known. We felt fortunate that the producers were a local telephone call away, and that we have succeeded in building bridges between ICPSR and the library. In similar undertakings, it would be helpful if producer and describer coordination were formalized at the outset instead of being developed ad hoc along the way.

Educating users. Providing service for computer files is more demanding than for traditional materials. As the project progressed, we began to ask ourselves whether bibliographic systems should go so far as to supply help screens for the user. Clearly, users need assistance in using computer files, but is self-help appropriate for materials of this complexity? How do we as describers and service providers indicate to potential users the difficulty of manipulating this material and the expertise necessary to do it successfully?

The Benefits

The ICPSR-in-RLIN project yielded a number of benefits to the University of Michigan Library. It provided an impetus for cataloging materials not only new to the library but also not actually under its roof. It opened up avenues of communication between the library and ICPSR. It provided improved access to an important and widely used set of resources both at the University of Michigan and the research community beyond. We hope that other research libraries can learn from our experience and will attempt similar cooperative ventures to catalog computer files. When they do, the entire research community will benefit. ■



Computer Files in the Library: Training Issues

By Katherine Chiang

At some point, librarians facing the challenge of incorporating electronically stored information into their library realize that information stored electronically is not a flash in the pan but the first twinkle from a supernova. That realization has come to most librarians and they are asking: Where are my sunglasses? Or, how do we cope with computer files? A big part of this question is, what do we have to know to provide patrons with access to information in electronically readable formats, and how do we find it out?

Integrating computer files into the collections and services of the library requires the selection, acquisition, and cataloging of files, and beyond that, providing the services necessary to locate and use the files: circulation, reference, and instruction. Each of these tasks requires certain skills—the ability to do something—and expertise—the possession of the information a skilled person needs to do things well. Some skills and expertise are already within the repertoire of the librarian, but some are not. A number of factors associated with each task influence when training in computer file services can take place and how it can be accomplished.

Selection

It takes tenacity and creativity to track down sources of computer files, whether they are mainstream commercial publishers, less obvious sources (data files advertised to practitioners in specific fields or professions), or obscure sources (files not for sale, although the producer is willing to duplicate and distribute them). Tracking down obscure sources is generally done in response to a specific request from a patron. It requires the same skills now used to find "gray" literature, where in many cases the most valuable selection tool is the telephone.

In addition, selectors must know enough about the advantages and disadvantages of the different formats on which computer files are distributed (optical discs, tape, floppy disks, online access, maybe even punched cards) to evaluate which would be most useful for their institution. Selectors must be able to take costs, hardware requirements, implications for service, projected use, and patron skills into account. (The last punch card reader at Cornell was discarded in 1989.) They must be able to order the proper version within a specific format for their patrons' or institution's hardware.

Acquisition

Before acquiring files, it is often necessary to contact the vendor to verify that the data are available in the appropriate format. Staff must know enough about computer files to ask the right questions about disk size, operating system, machine memory requirements, tape density, and labels.

Frequently, orders for tapes must be prepaid and the library must send a blank tape. Staff must know how tapes and disks are labeled or otherwise identified and how to handle the media they receive.

Katherine Chiang is computer files librarian at the Mann Library of Cornell University.

Traditionally, acquisitions staff verifies that incoming books and journals conform with what was ordered. Checking information that is stored electronically is a more complicated task requiring staff to know how to install software, load tapes, and "read" some of the stored information to verify that the data received are the data ordered and that all data ordered are present with no missing or garbled records. They must create backups of the data and may even need to reformat data for local storage requirements, for example, by reformatting a file into an SAS data set.

Cataloging

Cataloging—the identification and description systems used to locate relevant files—is a crucial component in creating a computer file service. The level of cataloging a library performs depends on how the data are to be made available.

If only library staff access the file, an idiosyncratic system designed for internal use might suffice. If the files and records are to be shared among libraries, use of the MARC format for bibliographic description is necessary. But for a truly useful description of a file from the patron's viewpoint, a MARC record has significant drawbacks. Because data can be easily moved from floppy disks to tapes, or compact disks to floppy disks, the bibliographic description tied to the container can be rendered invalid within seconds. The cataloger must supplement the minimal subject description allowed by Library of Congress subject headings with a high level of content description that can travel with the file through any format transformation.

Adequate subject description has always been a problem for libraries (making abstracting and indexing services a necessary supplement to bibliographic descriptions). With the decreases in computer storage costs and the rise in electronic versions of data documentation, the idea of providing detailed descriptions of file contents along with the bibliographic description of a file has become a realistic ambition.

In 1990, the creation of such descriptions requires that someone with technical expertise create supplementary databases with this information or modify existing databases such as NEDRES (National Environmental Data Referral Service) for local application. In the future, national shared databases such as RLIN® may offer file descriptions that go beyond a broad description of the contents of a file and offer much more detail than traditional subject headings would provide. (For instance, in the case of a social science data set, this could include information on variables included in the data, e.g., age, race, and religion.) Catalogers will need to know how to transfer documentation and variable lists from the computer file to be cataloged into the record format of the shared database.

Service Implications

The range of possible services for computer files is wide. A library can confine itself to directing users to a file. Beyond that, it can circulate, duplicate, make subsets of a file, and, ultimately, instruct patrons in computer file use.

Infrastructure. The options for services in any given library are determined by where the supporting hardware is. If a library does not own the mainframe needed to access a computer file on magnetic tape, it could buy the tape and send it to the computer center for service and maintenance. Reference staff could then direct users to files in the computer center, and either the staff at the computer center or the patron would be responsible for extracting the data of interest. Or, the library could keep the tapes and lend them to patrons, leaving it to the patron to find a

machine capable of reading and retrieving the data (much as we expect patrons to handle print information).

If some of the collection is on compact disks, the library could provide a compact disk workstation for each disk it owns. It could provide compact disk readers configured with the software to run each product and check the disks out, to be used on the library machines—much as it does microfiche. Or, as patrons buy compact disk readers for their own microcomputers, the disks could circulate outside the library.

If a library intends to provide the total infrastructure, from mounting a tape on the computer to data extraction, it will need a full range of technical and hardware resources. The tapes and disks must be stored, maintained, exercised, or recopied for preservation or circulation. Compact disks must be cleaned regularly. These housekeeping tasks are routine activities for computer centers and data archives and can be adopted by the library. To do so, library staff will also need experience running and maintaining mainframes or minicomputers, microcomputers, and peripherals such as printers, compact disk readers, and modems. They must know operating software, a statistical package or other program to handle data extractions, and possibly database management, text processing, communications (for access to in-house, local, and national networks), and even graphics or spatial database programs.

Mediation. Library staff can refer patrons to the location of files and documentation, consult with patrons on the types of extractions and manipulations possible for a particular file, help patrons extract subsets from files, and extract subsets from files for patrons. Librarians can also provide instruction on how to use data files.

To offer these services, librarians must know which indexes or finding aids contain information about computer files. They need to understand how documentation is organized. If the library is extracting data for patrons or advising on data manipulation, someone on the staff needs the requisite technical expertise, which can include knowledge of statistics and statistical computing packages, familiarity with the data set, and overall experience in general computing.

Skills and Expertise

In summary, the new skills and types of expertise required for librarians to carry out the tasks and services described above include:

- Ability to use finding aids (e.g., catalogs, documentation) to determine if a computer file can answer a patron's request.
- Knowledge of new sources to search for appropriate computer files.
- Knowledge of data on storage media and ways they can be written and stored (e.g., ASCII or SAS file).
- Ability to instruct patrons in the use of computer files.
- Knowledge of issues of storage media and machine compatibilities.
- Experience in creating databases and extracting, manipulating, or reformatting data.
- Knowledge of database design and construction (i.e., building one and determining what to put in it).
- Experience in system, tape, and machine maintenance.

Training

How do library staff acquire these skills? If libraries cannot hire people with the needed skills, they must train existing personnel. Several factors affect who is trained, how they are trained, and in what order. These include how the service is to be provided, and who is providing the service, their abilities, and their learning styles.

Level of service. The number of people to be trained is tied to the level of convenience offered. If the service is provided to patrons on demand, the library will need a full complement of staff able to provide that service. If access to computer files is limited to scheduled hours, fewer trained staff are required. If service is limited to appointment only, as few as one or two staff will probably be adequate.

Structure. The way the service is structured will influence which new skills must be learned by any one person. If the service is oriented vertically, so that one person or department is responsible for providing access to computer files, that person or group needs to know everything, from where to find files to how to extract data. If the structure is horizontal, and a number of individuals or departments are each responsible for providing one aspect of service, then the appropriate computer files skills are added to each department (e.g., acquisition, cataloging, and reference). Each approach has its advantages and is suitable for different situations. But the choice affects staff and what they will be required to learn.

Concept maturity and staff abilities. The difficulty of providing a new service is determined to a significant extent by the maturity of a service both in a specific library and in librarianship at large. The automation of libraries provides a good example of how the maturity of a service affects the resources required to provide it—including staff training. Twenty years ago libraries spent large amounts of resources installing automation systems for the first time. Because the systems were completely new, libraries were forced to establish procedures ad hoc. Now they can buy third-generation turnkey operations complete with staff training manuals. Integration of computer file service will follow the same pattern: early efforts will involve a great deal of trial and error, but with experience the service will become routine.

An additional element in training is that some personalities are more able to deal with nascent services than others. It is possible to identify several types.

The first person (or library) to do something is the innovator. The probability of false starts, dead ends, and time-consuming mistakes is highest for trailblazers. If a library has staff who are able to work independent of tradition, to synthesize, and to pair needs with resources creatively, that library will have an easier time creating services from scratch.

The adapter can work from another institution's model to synthesize a local version. Adapters must understand both the local situation and the overall concept of the service well enough to tailor the model to their own institution.

Most staff are able to learn a new skill or develop expertise once the systems have been established in the institution—these are the mainstream types.

Finally, some staff will always be hesitant about developing new skills or expertise. They require cookbook procedures, and learn those procedures rather than the underlying principles of the service. Some may never understand the underlying principles, or only reach this understanding very slowly.

Staff abilities also vary in other ways. Some people find it difficult to speak in public—others have difficulty working with computers. In planning how to move ahead with computer files, libraries will need to evaluate their staff, determine where their strengths and weaknesses lie, and find the pace they are best suited to.

Learning styles and options. Some delivery methods work better than others in imparting skills and expertise. Also, learning styles differ from one person to the next. Some learn from reading or from lectures alone, others do better seeing demonstrations, while others must have practical experience with an activity to comprehend it.

Since services generally require that staff have a combination of theoretical background and applied skills, most people will need a combination of methods. Reformatting a data set requires a knowledge of rudimentary programming (best taught by lecture), plus practical knowledge of how to log on to the computer and run a program (best taught by demonstration or hands-on experience).

Training can be accomplished in a number of ways: workshops, seminars, formal classes, demonstrations, manuals, apprenticeships, video/media presentations, workbooks, computer-assisted instruction, and one-on-one. Local training options will vary. Colleges and universities generally offer formal ways to acquire technical skills. Does the computer center offer classes in mainframe tape transactions? Is there a class on using SAS offered nearby? Are there existing training programs for similar services? Data archive training will give a librarian many of the skills he or she needs to run a computer file service in a library.

Each library will have to determine what is needed and who should acquire the new skills and expertise in light of issues previously outlined and its staff's existing skills and expertise. For some functions the required skills are not new—but the background expertise is (e.g., selection of computer files). In other cases skills are new and used against existing background expertise (e.g., technical knowledge.) Identifying training needs and developing systematic training are crucial to the effective and efficient addition of computer file services to the library.

In an ideal world, an institution would have the luxury of structuring its training program to offer a panoply of options along a carefully designed time line. In reality, training librarians to provide computer file services will be a series of compromises of unknown duration. But with organization and commitment, libraries can train their staff to offer access to computer files. Indeed, they will have to. ■

Results of the Research Libraries Group, Inc., Workshop on Machine-Readable Data Files

**Held at Stanford University,
September 28-29, 1989**

Agenda

Thursday, September 28

Morning:

Collection Management and Development Issues
Margaret Johnson, University of Minnesota

Small group discussions

Full group reconvenes: reports and discussion
Facilitator: Joseph A. Boisse, University of California at Santa Barbara

Afternoon:

Bibliographic Control and Access Issues
Lynn Marko, University of Michigan

Small group discussions

Full group reconvenes: reports and discussion
Facilitator: Carol Mandel, Columbia University

Friday, September 29

Morning:

Public Service Issues
Anita Lowry, Columbia University
Katherine Chiang, Cornell University

Small group discussions

Full group reconvenes: reports and discussions
Facilitator: Joan Chambers, Colorado State University

Afternoon:

Final wrap-up
Facilitator: Joan Chambers

Workshop Attendees

Ms. Patricia Vanderberg
Associate Librarian
212 Library
University of California
Berkeley, CA 94720

Mr. Joseph Boisse
University Librarian
3589 Library
University of California
Santa Barbara, CA 93106

Ms. Nancy Evans
Data and Information Services Manager
Carnegie Mellon University
Warner Hall 304
5000 Forbes Street
Pittsburgh, PA 15213

Ms. Joan Chambers
Director of Libraries
Colorado State University
Fort Collins, CO 80523

Ms. Anita Lowry
Deputy Head of Reference Department
325 Butler Library
Columbia University
New York, NY 10027

Ms. Carol Mandel
Director, Technical Services Group
Columbia University Libraries
535 W. 114th Street
New York, NY 10027

Ms. Katherine Chiang
Computer Files Librarian
Mann Library
Cornell University
Ithaca, NY 14853

Mr. John James
Director/Collection Development and Bibliographic Control
Dartmouth College Library
Hanover, NH 03755

Ms. Phyllis Jaynes
Director of User Services
Baker Memorial Library
Dartmouth College
Hanover, NH 03755

Mr. Donald Kelsey
Preservation Officer
University of Minnesota Libraries
309 S. 19th Avenue
Minneapolis, MN 55455

Ms. Margaret Adams
Archives Specialist
Center for Electronic Records
National Archives and Records Administration
Washington, DC 20408

Mr. Paul Evan Peters
Library Systems Coordinator
The New York Public Library
Fifth Avenue and 42nd Street
New York, NY 10018

Mr. Rodney Phillips
Associate Director/Public Services
The New York Public Library
Fifth Avenue and 42nd Street
New York, NY 10018

Ms. Susan Kallenbach
Director, Technical and Automation Services
Bobst Library
New York University
70 Washington Square South
New York, NY 10012

Mr. Brian Nielsen
Assistant University Librarian
Northwestern University
1935 Sheridan Road
Evanston, IL 60208

Ms. Emily Fayen
Assistant Director of Libraries
University of Pennsylvania
3420 Walnut Street
Philadelphia, PA 19104

Ms. Bridget Winstanley
Chief Research Officer,
Information Services
ESRC Data Archive
University of Essex
Colchester, Essex CO4 3SQ
England

Mr. Erich Kesse
Preservation Officer
210 Library East
University of Florida
Gainesville, FL 32611

Mr. R. Max Willocks
Associate Director/Science and
Technical Services
210 Library West
University of Florida
Gainesville, FL 32611

Mr. Dale Flecker
Associate Director for Planning
and Systems
Widener Library
Harvard University
Cambridge, MA 02138

Mr. Jerome Clubb
Executive Director
Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, MI 48106

Ms. Suzanne Thorin
Acting Chief, General Reading
Room Division
Library of Congress
Washington, DC 20540

Ms. Lynn Marko
Head, Monograph Cataloging
Division
Hatcher Graduate Library
University of Michigan
Ann Arbor, MI 48109

Ms. Margaret Johnson
Assistant Director, St. Paul Campus
Libraries
260 St. Paul Campus Central
Library
University of Minnesota
1984 Buford Avenue
St. Paul, MN 55108

Ms. Judith Rowe
Manager of Research Services
Computing and Information
Technology
Princeton University
87 Prospect Avenue
Princeton, NJ 08544

Ms. Marianne Gaunt
Associate University Librarian
Rutgers University Libraries
169 College Avenue
New Brunswick, NJ 08903

Ms. Tia Gozzi
Director of Technical Services
Green Library
Stanford University
Stanford, CA 94305

Ms. Barbara von Wahlde
Associate Vice President for
University Libraries
Capen Hall
State University of New York
Buffalo, NY 14260

Mr. Michael Keller
Associate University Librarian for
Collection Development
Yale University Libraries
P.O. Box 1603A Yale Station
New Haven, CT 06520

RLG Staff:

Jim Coleman, Program Officer
John Eilts, Program Officer
Connie Gould, Program Officer
Jennifer Hartzell, Director,
Corporate Communications
Leslie Hume, Director, Programs
Patti McClung, Associate Director,
Programs
Marilyn Roche, Senior Program
Officer

Workshop Discussion Results

Collection Management and Development

1. Participants agreed that the most urgent task is to determine more precisely the scope of the universe of computer files, both published and unpublished.

Cooperative efforts in collection development will be essential, but cannot proceed without a clearer understanding of what the collection possibilities are. The group observed that we are fortunate to witness in computer files the creation of a new format, and have an opportunity to develop practices for their management as the format takes hold.

Although the existing MDF file in RLIN® is ultimately the logical place for assembling information about individual files, the group agreed that other tools might be appropriate for capturing related information (e.g., on user documentation and code books), particularly while the format is still developing. The RLG Conspectus® is one possibility for providing high-level collection description. Some advocated a cooperative effort to collect basic information on as wide a range of computer files as possible in an inventory or directory; possibly the MDF file could serve as a home for the results of this effort.

Possible RLG role:

- **Lead efforts to identify published and unpublished computer files and their sources.**
- **Use RLIN® MDF and the RLG Conspectus® to provide information about computer files; consider the development of other tools if necessary.**

2. The development of standards or guidelines to assist selectors is essential.

No methodology for the selection of computer files now exists; selection philosophies also vary. Many believe that the ideal situation is integrated, discipline-driven, format-independent collection. This is easier said than done, for several reasons: funding is often separate from that of other formats; the wide variations in format (e.g., CD-ROM, diskette, tape) and type of file (e.g., bibliographic, textual, numeric) complicate selection; and libraries need more information about what users want.

The group unanimously endorsed the need for general computer file "consciousness raising," including training in computer file selection procedures. Computer file selection and service issues are inseparable; consequently, they must be considered in tandem, and awareness of their interdependence emphasized.

Possible RLG role:

- **Sponsor general educational workshops and training in the selection of computer files, both published and unpublished.**
- **Develop guidelines for determining the costs and benefits of different means of access to computer files (e.g., mainframe, remote, offline).**

3. Unpublished computer files are especially difficult to collect, because they are produced outside the normal channels. Further, they often lack features associated with the publication process such as documentation and bibliographic access.

Libraries are often unwilling to collect or provide access to computer files created by their own faculty, no matter how significant, because they often lack documentation and are not in final form. Although most institutions are not currently interested in becoming computer file publishers, some expressed interest in the creation of ICPSR-like centers or data archives to manage the explosion of unpublished data resources. (ICPSR: The Inter-university Consortium for Political and Social Research, a central repository for machine-readable social, economic, and political data.)

Possible RLG role:

- **Investigate the creation of "value-added" enhancements for unpublished computer files through pilot projects.**
- **Establish formal ties with data centers and archives to facilitate the intellectual management and sharing of computer files.**
- **Stimulate the formation of one or more ICPSR-like centers.**

4. As more information is available only in electronic formats, preservation of computer files will become increasingly important.

In particular, there are critical issues involved in the long-term preservation of information produced on an unstable medium (e.g., magnetic tape). In addition, there is a host of problems relating to file formats that become outmoded or are only usable on obsolete equipment. Although reformatting is clearly a potential solution, the policies and procedures for preserving computer files need to be addressed.

Discussion also underlined the importance of selection decisions in determining what computer files are preserved, how they are preserved, and for whom.

Possible RLG role:

- **Identify preservation issues within a wide framework and begin to prepare for future needs.**
- **Investigate reformatting questions.**

Bibliographic Control and Access

1. Traditional notions of bibliographic control and its fundamental purpose with respect to this format need to be re-evaluated.

Is the primary function of the bibliographic record to communicate information about computer files? Some participants questioned whether this interpretation is too limited. Whatever the final outcome, a consensus within the research library community is needed. Issues 2, 3, and 4 reflect this need for re-evaluation.

Possible RLG role:

- **Provide a forum for these discussions and leadership in re-evaluation.**

2. The RLG membership must work toward increased participation in building a shared database of information on computer files.

Two conflicting requirements exist: the need for more breadth of cataloging (i.e., the need for as many computer files as possible to be catalogued) and the need for more depth in cataloging than many existing records provide. Realistic cataloging expectations can help to answer the first need. Should an inventory or directory of computer files be established—a "computer file awareness file"? Many believe that the RLIN® MDF file can serve this purpose through the addition of acquisition-like records. Many institutions might be willing and able to contribute information at this level, and the purpose of the MDF file could thus be expanded.

The group also agreed that cataloging could be advanced through encouraging the practice of cataloging in publication among computer file producers of all types.

Possible RLG role:

- **Encourage its member institutions and others using RLIN/MDF to expand cataloging through the use of acquisition-like records.**
- **Develop guidelines for cataloging in publication and promulgate them among computer file producers.**

3. Expertise in computer file cataloging is scarce. Also, existing cataloging practice, as reflected in MARC and AACR2, needs to be reviewed.

Additionally, computer files issued in multiple versions pose difficulties as yet unresolved.

Possible RLG role:

- **Sponsor training workshops for computer file catalogers.**
- **Provide a forum for the intellectual issues connected with computer file cataloging.**

4. The fixed terminology that now exists for describing computer files is not adequate.

Some believe that an expanded thesaurus would work; others question whether any fixed vocabulary can do the job. Yet the use of uncontrolled keywords to describe computer files also presents problems.

Possible RLG role:

- **Address the question of how to provide satisfactory intellectual access to computer files. If necessary, RLG can also enhance the MDF file, e.g., by providing thesauri or indexes.**

5. A critical need exists for more analytics in the cataloging of computer files.

In particular, information about individual texts in large text collections on optical discs is now difficult to obtain. Guidelines for cataloging and standards for the description of files also need to be developed.

Possible RLG role:

- **Develop cooperative projects to produce analytic cataloging for large computer file collections.**

6. The issue of ownership is fundamental.

Some libraries are reluctant to catalog files they do not own, since they cannot be responsible for the maintenance of or public access to such files. Other libraries, however, are willing to provide information about files not controlled by the library. This issue will continue to present a challenge since computer files are also collected by units outside the library—departments, computer centers, etc.—on many campuses. Although this is clearly a local choice, further discussion seems warranted.

Possible RLG role:

- **Provide a forum for this issue.**

Public Services

1. A think tank (i.e., a forum focusing on computer file issues) is necessary to help libraries better manage the impact of technological change brought about by the growth of computer files.

Experts representing libraries, scholarly and professional associations, data archives/repositories, vendors, computer centers, library education institutions, and computer file producers should explore such issues as:

- Appropriate technologies for various applications
- Changing technologies
- Multiple access to single sources
- Equipment standards
- Documentation standards
- Role of library/role of computer center

in a forum conducive to addressing computer file issues and concerns.

Possible RLG role:

- **Play a prominent role in convening such a forum.**

2. Consciousness raising and subsequent training are major issues as librarians are increasingly required to develop computer expertise.

Staff, faculty, scholars, academic administrators, creators of computer files, vendors, and library educators need to increase their awareness of the information technology environment. They should be trained in the use of data files and provided reasonable access to files and the equipment needed to use them.

Possible RLG role:

- **Function as a clearinghouse for sharing the training materials developed by individual institutions as well as for coordinating the cooperative development of materials.**
- **Design and develop apprenticeship or internship programs and offer training for the consortium through workshops and conferences. Such programmatic planning for the consortium may serve as an impetus for other advances in computer file policy and practice.**

3. Allocation of resources becomes a prime concern for libraries with the increasing availability of computer files.

Determining the portions of the budget to be allocated to print materials versus database and software components for various subject disciplines is difficult. In addition, libraries may need to restructure collection budgets to accommodate the acquisition of alternative formats and equipment necessary to access the new sources of data. The uneven enthusiasm for computer files and consequent uneven allocation of resources among disciplines further muddies the waters.

Possible RLG role:

- **Direct consortial attention to the tradeoffs that result from advances in technology and to the impact of resource reallocation on the accessibility of research information.**
- **Assess computer file needs across the consortium, and determine how the group can work together most effectively to meet them.**
- **Negotiate with vendors to obtain group rates on computer file products and equipment.**

4. With increased enthusiasm for sharing computer files, the need for a union catalog of computer files is greater than ever. Better knowledge of resources within member institutions is also a critical need.

Both intracampus cooperation, among departments, libraries, computer centers, data centers, and research archives to produce, collect and/or service computer files, as well as interinstitutional cooperation among libraries, scholarly societies, other academic associations, and industry are necessary to facilitate identification, evaluation, and sharing of computer files—especially unpublished files. Libraries retain the primary responsibility for the sharing function.

Possible RLG role:

- Be a catalyst for efforts to extend knowledge of local computer file resources and expand the union catalog.
- Facilitate communications, influence the development of grant support for computer file collections and equipment, and develop a central directory with information on the availability and accessibility of computer files.
- Mount selected databases on RLIN.

5. *Copyright guidelines should be clarified.*

Possible RLG role:

- Represent the consortium in initiatives to influence licensing agreements and copyright guidelines that would maximize access to computer files while protecting legitimate copyright claims.

Appendix

Project Summaries from
Managing a New Library Resource: Results of the Machine-Readable Data File Project in Six Member Libraries
(Mountain View, CA: The Research Libraries Group, Inc., 1989)

Computer Files in the Research Library: Issues of Collection and Access

Cornell University

The selection, acquisition, cataloging, processing, and access to computer files should be brought into the mainstream of the operations of the research library. Given that premise, what are the skills library staff must have in order to provide access to computer files? The major issues affecting bringing computer files into the mainstream of the library include the following. Sources (federal government, value-added commercial, private, state) affect the ease with which a library can incorporate the data into its collection. Selection from many sources is difficult and extraction of data from all but the commercial sources is also difficult. The different formats in which a given data set might appear can make selection challenging. The formats also affect the processing and cataloging of the files.

The skills a staff must have can be divided into four areas: hardware/technical, software/programming, data issues, and reference.

Hardware/technical: Day-to-day requirements of the maintenance of tapes, mainframe housekeeping to keep track of the collection, mechanics of running machines (both micro and mainframe), networking of machines, data transfer from machine to machine.

Software/programming: Communications programs and protocols, mainframe batch and interactive operating systems, database construction programs, statistical programs, tape reformatting, data extraction.

Data issues: How data was collected for each specific file, error checking, best format for the data.

Reference: Where to find data, evaluate appropriate data format, describe services to patrons, conduct reference interview.

Services for	Hardware Requirements	Ability to/Skills Needed
Commercially produced remote access online systems	Dumb terminal or microcomputer with communications software and telecommunications capability	Use a standard communications applications program/ training in the use of the online system
Commercially produced floppy and compact disks	Microcomputers with compatible operating systems and compact disk readers	Use a microcomputer, its operating system and learn the retrieval software provided with the product
Magnetic tapes of data in simple table format	Mainframe access, software to subset data	Use a mainframe job control language/basic knowledge of a statistical software program

Services for	Hardware Requirements	Ability to/Skills Needed
Magnetic tapes of data	Mainframe access, software to manipulate data	Create and execute mainframe programs
Locally produced online systems	Microcomputers of mainframes with software to create interactive systems	Create and execute complex mainframe systems
Network, remote access, to locally produced systems	Microcomputers or mainframes and software for remote access	Network complex systems

Conclusions

- Selecting computer files is difficult because the same data can be distributed by several companies or publishers.
- Acquiring computer files is not straightforward. The complications presented by format, active files, and processing variations are a challenge to the staff.
- Cataloging computer files to national standards is also a challenge because cataloging rules were not redesigned, but are only modifications of the rules for cataloging print material.
- Variable-level access to the data collection is essential.
- A library should select a level of collection and services that it can support adequately, build its systems and audience, and then upgrade its collection and services, if necessary.

Capturing Information about Machine-Readable Files in the Humanities

Dartmouth College

The goals of the project were to test a survey method and a data collection/inventory form, examine issues of bibliographic control, continue the development of Dartmouth's MRDF database, and consider the role of the library as a provider of access to information resources not necessarily owned by the library.

Findings/Achievements

Data collection: Personal interviews of faculty and other data owners are much more effective than mailing survey forms. A data collection form can be designed that is clear and brief, collects sufficient information for cataloging purposes, and can be used by faculty with various levels of technical expertise.

Bibliographic control: Dartmouth's approach to local bibliographic control is to maintain a separate MRDF file as part of the online catalog using a simple cataloging format based on, but not as detailed as, the MARC DF format. Following our experience during the project, there is evidence that a MARC-type database may be more desirable.

The MRDF resources identified during the project were entered into RLIN® using the RLIN® MDF/MARC DF format. Several information elements not part of the RLIN® format needed to be added to make the records useful.

Dartmouth MRDF database: Records for the MRDF identified during the project were entered in the database and issues of updating and evaluation of the database were identified.

Library ownership of data: Dartmouth's approach was to include records for both library-owned and non-owned MRDF. The library's mission is to provide access to information resources no matter where they may be located. It is, then, appropriate to include information in the library's catalog about MRDF held by faculty.

Issues Needing More Attention

- Catalog format revisions
- Sight-unseen cataloging
- Ownership of resources represented in the local catalog and RLIN®

Progress in MRDF: The Machine-Readable Data File Project

University of Florida

The goals of the project were:

- To enhance our MRDF services by ascertaining user needs through a survey and to prepare a model collection development/management policy statement, leading to the organization of a MRDF collection which meets the needs of the University of Florida (UF).
- To catalog a significant portion of data sets currently owned by UF into the online catalog, RLIN®, and OCLC.
- To document procedures for technical analysis and physical care of the data set collection.
- To design a service policy that incorporates copyright, contractual, and data integrity constraints and to develop appropriate procedures to implement the policy.
- To define technical, source, and use potential parameters for data to be added to the collections of UF and to obtain data sets now held in faculty offices for the library collection.

Work on the first goal led us to the conclusion that MRDF do not require a separate collection development policy but will much more appropriately be incorporated into the overall collection development policy of the library. Decisions to purchase MRDF require them to meet the same criteria as other library materials. However, even in a library that has had many years of experience with MRDF, it is very difficult to get bibliographers and selectors to include this material in their normal work flow.

The study affirmed our policy that the library should collect MRDF and house them in the library. It is then necessary to adequately represent these materials in the library's catalogs. We were especially happy that the grant enabled us to catalog a large number of files and put them into our own online public-access catalog (OPAC). This has made our collection more accessible to UF faculty and students and has also enabled researchers outside the university to become aware of these holdings (e.g., at least one Ph.D. dissertation being written at another university relies on one of our Brazilian data files).

We have succeeded in developing and documenting procedures for technical analysis and for the physical care of data files. We still need to further develop our public awareness program. Librarians should be just as familiar with MRDF and just as ready to serve the users with this format as any other. The grant enabled us to define and clarify our service policy.

The project confirmed our decision that we would not catalog and put into our OPAC material (including MRDF) that we did not own.

Our biggest disappointment was that we were neither able to contact personally as many faculty members as we had anticipated nor obtain as many data files from faculty offices as we had predicted. This is an ongoing

project, however, and we are still receiving data files. We have just completed negotiations to receive a major LANDSAT data file and are in the midst of negotiating for another major data file. These are in addition to the data files being regularly purchased by bibliographers and selectors.

The survey effort confirmed that faculty hold many data files and that they are anxious for the library to be the repository of the data files. They like having data files cataloged and generally available but they are not going to give up their own data files very easily. It is also clear that almost all disciplines now generate data files and that if the library is going to continue to service them, it must broaden its perspective.

Policy Development for the Acquisition of Machine-Readable Data: A Demonstration Project

Northwestern University

Northwestern University Library, with assistance from the Northwestern Medical Library, Academic Computing and Network Services, and the Northwestern Transportation Center, participated in the RLG MRDF Project in order to develop a survey methodology for information on MRDF needs and uses among faculty at academic research universities. Because the objective was to begin the process of formulating university-wide MRDF policy, Northwestern felt it was important to examine user perspectives. Though Northwestern has a reputation in research library circles for being technologically advanced, in the matter of MRDF we are more or less typical, having a separately administered computing center carrying out what limited data archive functions exist for researchers (essentially ICPSR-related administration). At the time the project began, the library had had no formal involvement with MRDF.

The process of involving faculty in articulating policy objectives for acquiring machine-readable data revealed a variety of issues not often acknowledged within librarianship. Many of these issues make our concept of providing "information in whatever form" seem over-simple.

The following conclusions came out of our analysis of faculty interviews and surveys:

- Faculty do see a need for systematic training in this application area, for themselves as well as their students.
- There is currently insufficient information about data sets available locally and from other sources.
- Procedures and policies for arranging consulting services are at present unclear.
- The existing computing network is not adequate to meet research requirements, especially for accessing remote bibliographic files.
- Budgetary support for data acquisition is a necessity if use of machine-readable data is to grow significantly.
- Microcomputing must be seen as a significant platform for ultimate data use, and thus delivery mechanisms to support that platform will be required.
- Many data sets are suspect in regard to their quality or reliability, and thus the library must be cautious in advocating a program to acquire data sets. Acquisition alone is an insufficient primary goal.
- Compelling reasons exist for users not to share data on which they are currently working; thus a collection development approach which relies heavily on data sharing is unlikely to develop sufficient support.

- Current data set use is primarily in research rather than teaching, which makes collection development problematic: Do you serve specific researchers, who may already have funds and mechanisms for data access?
- The absence of clear standards for documentation of machine-readable data, in part as a result of the ease with which data can be reproduced for publication or subsequent use by others, means that libraries collecting MRDF may be saddled with responsibilities which in the case of print are assumed by publishers.

These conclusions have informed the ongoing policy development process at Northwestern, and it is hoped may encourage other research libraries, working closely with other stake holders in their institutions, to examine the broader issues.

"Covering the Bases" Executive Summary

New York University

The issues related to bibliographic access and reference services were discussed as Bobst Library planned to incorporate machine-readable data file cataloging data into BobCat, its online public catalog. Working together, staff from the Academic Computing Facility and the library developed a series of workshops, "Covering the Bases: Techniques for Data Research and Retrieval."

Enrollment in the course was open to a pool of interested graduate students and faculty following a universitywide publicity campaign. Of over 80 applications received, 29 people were enrolled for the five sessions.

Each two-hour workshop was taught by the coprincipal investigators and guest lecturers drawn from NYU library staff. Additional out-of-class time was made available for hands-on instruction in use of numerous computer systems and dial-up services, such as DIALOG, the Roper Center's POLL (Public Opinion Location Library) database, and ICPSR's Guide/Variables/Roll Call files.

The classroom used for the course was fully equipped with a video projection unit capable of handling computer output for demonstration purposes. Instruction of this nature is greatly facilitated by hands-on experience during the class session. The addition of four to six microcomputers would have allowed us to walk participants individually through log-on sequences, basic commands, and downloading capabilities before pursuing the exercises on their own. This type of lab setting would greatly facilitate future workshops of this type.

Course evaluation forms were distributed to all enrolled participants. "Personal interest" and "research needs" were cited by over 80% of the participants as their reasons for taking the course. Surprisingly, only 20% of the participants checked "teaching/advising purposes" as the reason they took the course.

Overall, participants were positive about the workshop. Course content received high marks. Participants were satisfied with the lecture/demonstration/question-and-answer teaching style as well. Several participants voiced complaints over the interdisciplinary nature of the course. Many of the faculty participants, in particular, required assistance on current research projects and expressed dissatisfaction with what they perceived as tangents in the course content.

Time was the major limiting factor. The five-session format did not allow for enough discussion or attention to individuals' ongoing research needs. As the number of data resources available to researchers increases, the need for continued training grows. Library staff must upgrade their professional skills, and cooperation with computer center staff must continue in order to provide the desired level of service for this valuable research material.

Providing Local Access to Machine-Readable Data Files: Choices and Tradeoffs

University of Pennsylvania

The University of Pennsylvania libraries participated in the RLG Machine-Readable Data Files Project to collect information for assessing the most cost-effective means of providing access to machine-readable data files. The choices are remote access via a time-shared service such as DIALOG or BRS, local CD-ROM, or local online access.

Several factors were identified that must be considered:

Timeliness: Are the data time-sensitive? Do users have a critical need for the latest information?

Anticipated use: Are there many potential users? Is use expected to be very heavy?

Simultaneous users: How many users need to be able to search the file at the same time?

Remote access: Do users need access to the data from various locations, e.g., classrooms, dormitories, laboratories, or home?

During FY 1988, users logged over 2,800 hours of online search time in the libraries. Of these, 2,161 hours were from users conducting their own searches and the remainder were for mediated searching. CD-ROM use during the study was very heavy as well, although comparable hourly use data is not available. Local database use now is over 30,000 hours per year.

The surprising conclusion from this project is that there is no one best method of providing access to machine-readable data files. Penn discovered during this project that all three forms of access—conventional remote time-sharing, CD-ROM, and locally mounted files—are needed. They serve different communities of users and meet different needs. Providing all three allows each to be used to the fullest and optimizes the tradeoffs.

The project also led to a number of unanswered questions: Could the libraries handle the extra traffic if access to local data files was available to the entire Penn community? What new databases should be loaded? Should the libraries press for networked CD-ROMs? Is mediated searching too expensive to continue? Finally, how can the libraries meet user needs for new and better services at a time when we are barely comfortable with the services we are providing now?